VOL. 13 No. 6, pp. 145-180 DECEMBER 1951

Philips Technical Review

DEALING WITH TECHNICAL PROBLEMS
RELATING TO THE PRODUCTS, PROCESSES AND INVESTIGATIONS OF
THE PHILIPS INDUSTRIES

EDITED BY THE RESEARCH LABORATORY OF N.V. PHILIPS' GLOEILAMPENFABRIEKEN, EINDHOVEN, NETHERLANDS

THE WIRE CAPACITOR AND OTHER COMPOSITE DRAWN PRODUCTS

by J. L. H. JONKER and P. W. HAAIJMAN. 621.319.4:621.396.6:621.771.31

The attempts that are being made to reduce more and more the dimensions of electrical components have led to the development of a new product, the wire capacitor, which not only has very small dimensions but possesses excellent electrical properties. This product is made by means of a special drawing process. Together with the magnetic material Ferroxcube this drawn capacitor makes it possible, for instance, to construct I.F. transformers which are considerably smaller than those hitherto employed.

The demand for small capacitors

The requirements to be met by electronic apparatus are increasing in number and becoming more stringent day by day. As a result a large number of components have to be incorporated in almost any apparatus, and yet at the same time the apparatus is required to be of limited dimensions. As examples we would mention the electronic computing machines, with their thousands of valves, relays, resistors, capacitors and inductors and the modern military electronic equipment, in which it is imperative that the dimensions and weight be minimized.

As a simpler case, let us consider a normal radio receiver. Generally speaking this contains 5 radio valves (some of which perform more than one function) and about 40 resistors, 20 inductors and 40 capacitors. For the dimensions of such an instrument to be kept within reasonable limits it is necessary that all these components should be made as small as possible. It is also of importance that various components can be combined into one small "unit", as for instance the I.F. band-pass filters, consisting of two inductors and two capacitors incorporated in a small can. Hence the components must sometimes match each other.

Much progress has already been made towards reducing the size of components. The resistors now being used are almost invariably small carbon resistors. The magnetic material Ferroxcube permits of smaller inductors being made. As far as capacitors are concerned, these are of various types. Paper capacitors, which are of rather large dimensions, are used in places where no particularly high requirements are to be met as regards constancy of the capacitance and suchlike. In cases where more stringent demands are made, the small but much more expensive mica capacitors are employed, which are of a better quality. In certain circuits (for higher frequencies) mica capacitors are often replaced by ceramic capacitors, which are much smaller than the other types, at least for the lower capacitances, but when such a ceramic capacitor is required to have a high capacitance and still be of very small dimensions this usually entails a larger temperature coefficient.

In addition to being small, it should also be possible for the components to be easily made within the tolerances required. Furthermore, they have often to answer stringent requirements of constancy in their properties during their life and must not be susceptible to the fluctuations in temperature taking place in any electrical apparatus.

The wire capacitor

The Philips' Works at Eindhoven have developed a capacitor of low capacitance (about 100 pF) which is smaller than any type hitherto known. This is the wire capacitor. In its most common form it consists of a small metal tube slightly less than 1 mm in diameter and about 5 cm long, in which is contained a metal core about half a millimetre thick, while the space between the core and the jacket is filled with a compressed insulating material with high dielectric constant. The total volume of such a capacitor is no more than 30 mm^3. This is to be compared with the 100 mm^3 of a ceramic capacitor and 1200 mm^3 of a mica capacitor of the same value and suitable for the same voltage (300 V). As regards dielectric losses, temperature coefficient, etc., the wire capacitor is comparable to the mica capacitor. Such drawn capacitor are shown in *fig. 1.*

The narrow space between the core and the jacket is filled with the insulating material in such a way as to make it highly uniform throughout. It would be impracticable to fill the capacitor as such. The basic material for the production of drawn capacitors is a metal tube 20 cm long, with outer diameter about 20 mm and a wall thickness of 2 mm. A wire core about 8 mm thick is inserted in the tube and centred with the aid of two rings acting as jigs. The annular space, which is quite wide enough for this to be done at this stage, is filled with an insulating material in powder form, which is firmly stamped in. The whole is then hammered and afterwards drawn out to a "wire" of about 40 metres in length and with the desired diameter of slightly less than 1 mm. Notwithstanding the great deformations to which the metal and the insulating material are subjected during this process, after the wire has been drawn the core and the insulation are still in place and a firm product is obtained.

The drawn wire is divided into pieces of the desired length, and at one end of each piece the jacket and insulation are removed, thus leaving the core bare. Connecting wires are then soldered onto the bared end of the core and onto the jacket. *Fig. 2* shows the product in the various stages of manufacture.

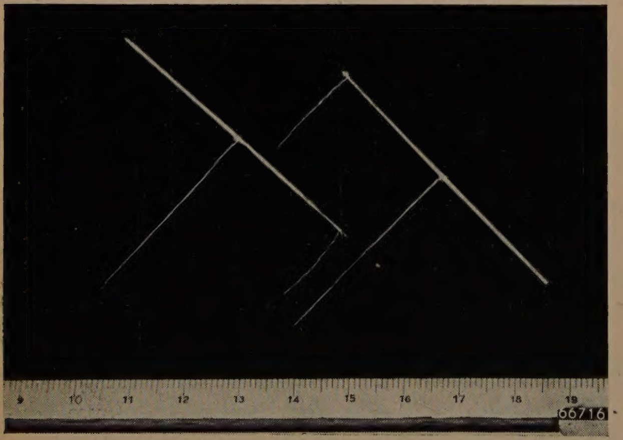


Fig. 1. Two wire capacitors.

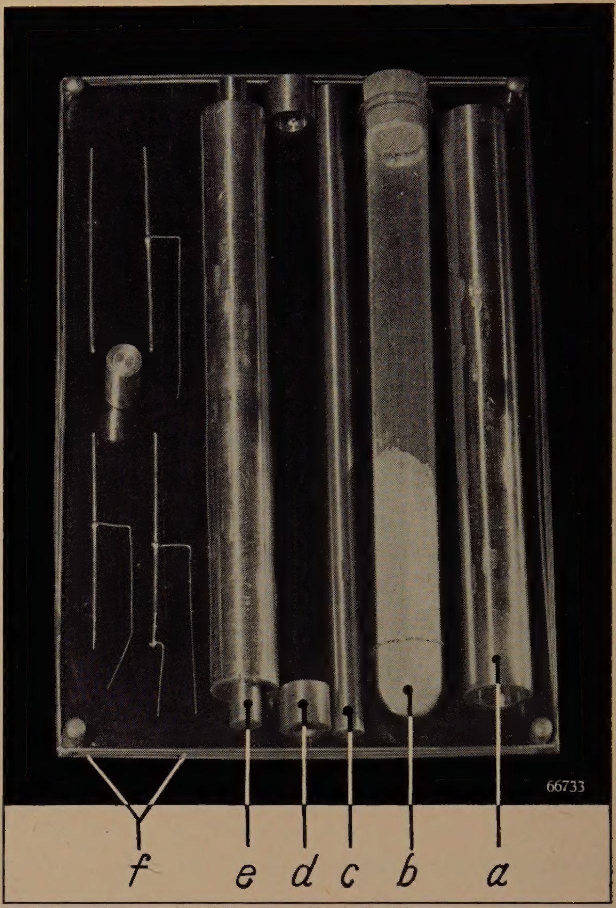


Fig. 2. Composition of the wire capacitor:
a) the jacket before drawing;
b) the dielectric with which the capacitor is filled;
c) the core prior to drawing;
d) jig rings for centering the core in the jacket prior to drawing (and after filling);
e) the composite system ready for drawing;
f) the drawn tube broken at the required length (upper left), with one connecting wire soldered on (upper right), the outer jacket partly removed by etching (lower left), with the second connecting wire soldered onto the core (lower right). Part of the 40-metres long, drawn "wire" is stretched round the whole.

In the development of this process several problems were encountered, which will be dealt with in this article. In conclusion some applications of the wire capacitor will be discussed.

In addition to capacitors, other products can also be made in the manner described here. We have in mind, for example, indirectly heated cathodes for electron valves, which likewise consist of two conductors (the filament and the cathode surface) separated by an insulating layer. Drawn cathodes can be used for various applications where normal cathodes fail, as will be shown. There are also other products of a similar construction which can be made by the drawing process.

Composition and manufacture

Materials

The materials from which the wire capacitor is built up have to satisfy certain demands. For instance, the metal of the outer wall and the inner conductor must lend itself readily to drawing, and at the same time it should show little tendency to oxidize, because oxidation would reduce the quality of the capacitor. As regards the dielectric the following conditions, among others, have to be satisfied:

1) It must have great permittivity (ε).
2) The dielectric losses must be low, tan δ (δ is the loss angle) should be less than 10^{-3}.
3) The temperature coefficients of ε and tan δ must be small; when used in electronic apparatus the capacitor is subject to temperature fluctuations, and these must not greatly affect the capacitance and the loss angle. This condition is already partly satisfied if the second requirement is fulfilled.
4) Only small differences are allowed between the thermal expansion coefficients of the dielectric and the metal, a point which will be reverted to later.
5) The insulating material must be easily pulverized; the grains may not be too hard, nor too large, as otherwise, during the process of drawing, they would scratch the surface of the metal wall and the inner conductor, thereby increasing the area liable to oxidation.

Extensive investigations have led to a combination of materials being found with which a capacitor can be made that has small losses (tan $\delta < 6{\cdot}10^{-4}$) and a very low temperature coefficient.

Details of manufacture

The 40-metre long "wire" obtained by drawing is stretched and divided into pieces of the length required. When dividing the wire there must not be any short-circuiting between the inner and outer conductors. There is no danger of this, however, since the jacket always bends somewhat outward when being broken. The insulating material is so firmly packed between the core and the jacket that there is no possibility of any of it falling out of the capacitor along the plane of rupture.

To bring about electrical contact with the inner conductor, the jacket and the insulation have to be removed over part of the length of the capacitor. This also affords an opportunity to give the capacitor exactly the desired value, by removing just as much of the jacket as is necessary to reach that capacitance. (The capacitors must therefore have a somewhat higher capacitance before part of the jacket is removed.) In practice the jacket is removed by electrolytic etching, with the jacket serving as anode. A hundred of the capacitors at a time are immersed in the etching bath to a depth corresponding to the length of jacket to be removed, and a voltage is then applied between all the jackets and an electrode placed in the bath. The current passing through the bath is continuously controlled; by the time that the metal of the jacket has dissolved the current has dropped to zero (or practically zero); the inner conductor of the capacitor cannot act as an electrode because it is insulated from the outer conductor. The insulating material is so firmly compressed during the drawing process that after the outer jacket has been etched off it does not of itself break away from the inner conductor, but has to be removed separately. After the etching the capacitors are washed in water in order to remove any traces of the etching medium, which might cause conductance between the inner and outer conductors. This treatment does not in any way impair the quality of the capacitors.

An entirely different method of removing the jacket is to shave it off with a lathe. For this purpose a special lathe has been designed, which is characterized by the fact that the cutter rotates, while the capacitor is clamped tight close up against the cutter; this is necessary on account of the natural lack of rigidity of the thin capacitor. The capacitor

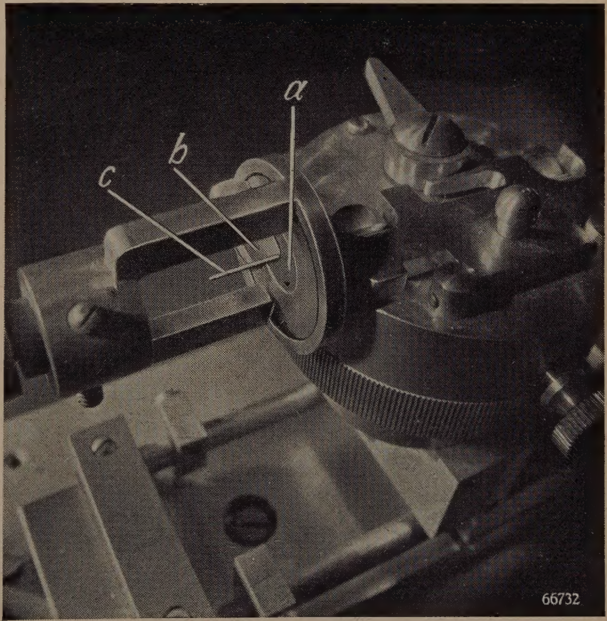


Fig. 3. Part of the lathe on which the jacket of the capacitor is removed over a length of about 1 cm. The capacitor is moved from right to left by the guiding die *a*. The rotating cutter (of diamond) *b* can just be seen behind the (turned) capacitor *c*. The automatic regulating device is not shown in this photograph.

is centred in the lathe by means of a diamond guiding die, while a diamond cutter shaves off the jacket over the length necessary to give the capacitor the value desired. The capacitance is measured while the jacket is being shaved off and as soon as it reaches a value within the required tolerances the rotating cutter is automatically stopped. By this shaving process it is possible to reach a tolerance of 0.2% in capacitance. By the etching method the degree of accuracy that can be reached is not so high, so that after this process the capacitors have to be sorted. *Fig. 3* shows a part of the lathe in question.

Physical properties of the wire capacitor

This is not the place to discuss fully the electrical properties of the wire capacitor, but we shall consider briefly two very important properties which determine the quality of this component and are largely influenced by the remarkable process of manufacture. These properties are the temperature coefficient of the capacitance, to be denoted here by a, and the losses in the capacitor, which are proportional to tan δ.

Temperature coefficient

Although every possible precaution is taken to ensure that the capacitor is filled as compactly as possible, air cavities between the grains of the powder are unavoidable. When the temperature of the capacitor is raised, both the jacket and the core as well as the grains expand. Since the expansion coefficients of the materials are not exactly equal, the "filling factor" of the capacitor changes with its temperature. The permittivity ε of the powder likewise changes with the temperature. Consequently the capacitance of the capacitor varies.

The materials, therefore, have to be so chosen that the temperature coefficient due to the difference in expansion coefficients is compensated as completely as possible by the temperature coefficient of the powder. Since, however, the materials have to satisfy many other requirements, complete compensation is very difficult to obtain. The temperature coefficient of the permittivity of the powder used is slightly negative. It appears that the powder expands somewhat more than the metal of the jacket and the core, so that upon the temperature being raised the filling factor is increased, which in itself leads to a positive temperature coefficient. For the resultant relative temperature coefficient of the capacitance, $a = 1/C \cdot \mathrm{d}C/\mathrm{d}T$, we find (at temperatures up to about 60 °C) roughly $+1 \times 10^{-4}$ per degree centigrade for a capacitor having a copper jacket and a copper core. A capacitor with a copper jacket and an aluminium core shows a higher value of a, viz. about $+5 \times 10^{-4}$ (see *fig. 4*). As fig. 4 shows, the curve representing $\Delta C/C$ as a function of T describes a loop: the variation in capacitance while heating up differs from that during the cooling of the capacitor. The curves also show maxima, especially those for the capacitor with copper core. No exact explanation has yet been found for these two phenomena; they no longer occur when the capacitor has been aged, in a manner presently to be discussed.

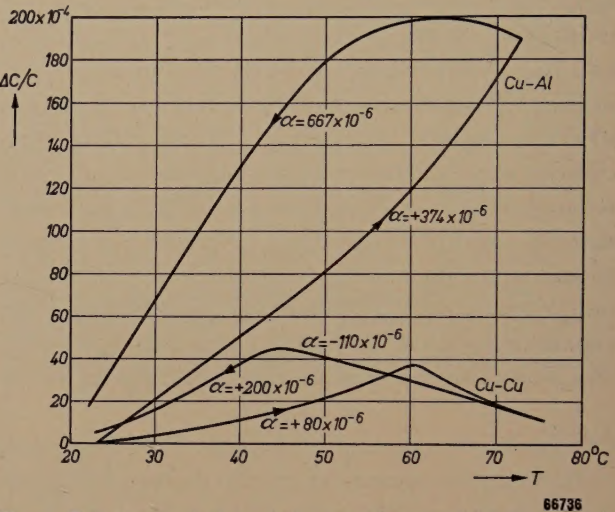


Fig. 4. Relative variation of the capacitance of a wire capacitor as a function of the temperature. No exact explanation has been found for the fact that the curve for increasing temperature differs from that for decreasing temperature. After the capacitor has been aged this "hysteresis loop" disappears. The two loops are respectively that of a capacitor with copper jacket and copper core and that of a capacitor with copper jacket and aluminium core. The latter shows a much greater temperature coefficient (slope of the curve). The values of the temperature coefficients are given in the drawing.

The temperature coefficient of the capacitance can be calculated if the expansion coefficients of the materials and the temperature coefficient of the permittivity are known. The results of calculations made agree well with the temperature coefficients as actually observed. Partly with the aid of these calculations it has been found possible to control the temperature coefficient within certain limits by varying the ratio of the dimensions of the jacket and the core.

Dielectric losses; ageing of the capacitor

The hammering and drawing processes cause the temperature of the capacitor tube to rise. Particularly during the hammering, when the temperature reaches about 80 °C the metal (copper) of the jacket and the inner conductor tends to oxidize, forming, for instance, Cu_2O, which is a semiconductor and thus causes losses. This oxidation is, of course, limited by the amount of air available inside the capacitor, which amounts to about 0.6 vol. %. The fact that oxidation is indeed the main cause of the losses is proved when a non-oxidizing metal (e.g. silver) is used for the jacket and the core; tan δ is then much smaller than the value of 6×10^{-4} mentioned in this article.

During the fluctuations in temperature, which regularly take place in practice, the capacitor alternately ejects and draws in air, so that the air inside the capacitor is partly renewed each time. Consequently it is to be expected that oxidation is started again time after time. And, indeed, tan δ appeared to increase, at least for some time after the capacitor had been made. Later, however, no further change in tan δ was to be observed, probably due to the whole of the surface being covered with a very thin layer of oxide. In order, therefore, to avoid trouble from an increase in tan δ during use, the capacitors are aged before delivery by keeping them for a few hours in an oven in which the temperature is alternately raised and lowered. The capacitors aged in this way have a constant tan δ also immediately after manufacture.

Applications of the wire capacitor

The wire capacitor has been specially developed for use in I.F. transformers. The principle of such a transformer is represented in *fig. 5*. It consist of two mutually magnetically coupled L-C circuits. The two coils L_1 and L_2 are each wound directly on a core of Ferroxcube, a ceramic magnetic material with high permeability [1]). The coupling is brought about by a rod of ferroxcube. The exceptional properties of ferroxcube, compared with the powder cores formerly used, allow of the coils being made so small that when mica capacitors are used the dimensions of the filter are almost solely determined by those of the capacitors. Thus it was desired to replace the mica capacitors by smaller ones, for example by ceramic capacitors. Now, however, the wire capacitor proves to be a still better solution for the I.F. transformer, since it is much smaller and, moreover, for the capacitances required it has a lower temperature coefficient than the smallest possible ceramic capacitor. *Fig. 6* is a drawing of a type 5730 I.F. transformer, cut open, in which two wire capacitors are employed; *fig.* 7 is a photograph showing the external appearance of these I.F. transformers. The dimensions of this cylindrical I.F. transformer are 60 mm × 27 mm, and the bandwidth for which it is suitable is 18,000 c/s at the (standardized) intermediate frequency of 452 kc/s. (I.F. transformers are also being made for

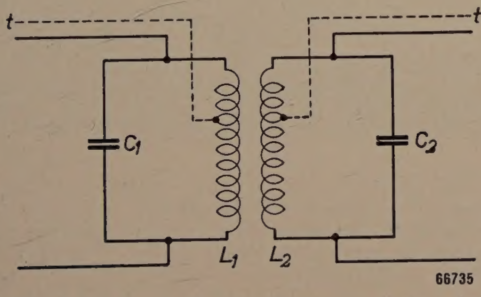


Fig. 5. Principle of an I. F. transformer consisting of two magnetically coupled L-C circuits. The tappings denoted by t can be used when it is desired to reduce the influence of resistances and capacitances outside the filter.

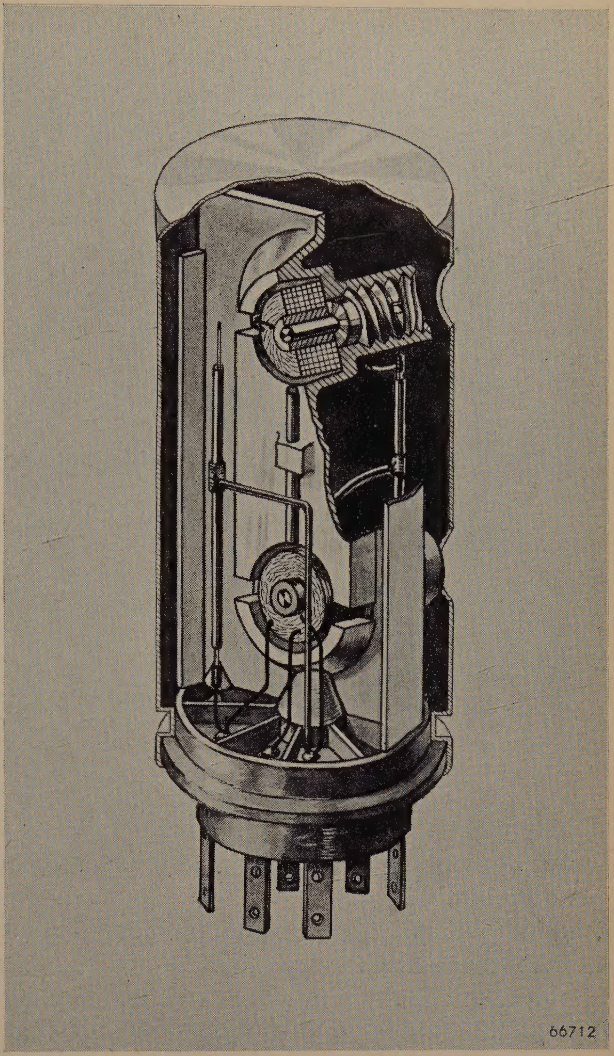


Fig. 6. Drawing of the I. F. transformer, type 5730, cut open. Note the ferroxcube cores, which can be screwed farther in or out of the coils by means of screws from the outside of the transformer. The wire capacitors can be seen on the left and right sides; the vertical rod in the middle is the ferroxcube coupling rod.

---

[1]) J. L. Snoek, Non-metallic magnetic material for high frequencies, Philips Techn. Rev. 8, 353-360, 1946.

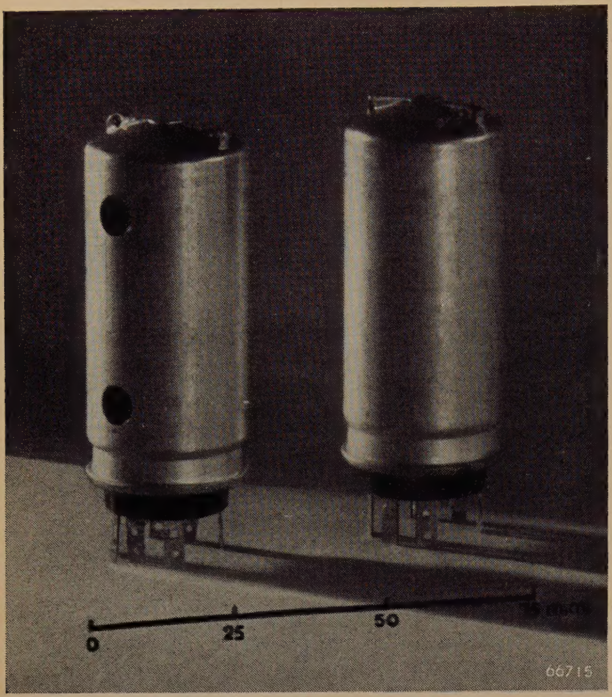


Fig. 7. The type 5730 I.F. transformer. The openings are made in the side for adjusting the screws with which the cores can be screwed in and out of the coils.

slightly different frequencies.) The circuits can be tuned by screwing the ferroxcube cores more or less deep into the coils, it being possible to do this from the outside. The temperature coefficient of the I.F. transformer is very small; for 1 °C variation in temperature the frequency drift is no more than 5 c/s.

In recent years Philips have brought onto the market a still smaller I.F. transformer of very similar properties, the dimensions of which are 36 mm × 25 mm × 10 mm. This has been made possible by employing somewhat shorter wire capacitors of the same value. Two of these "microtransformers" (type AP 1000) are illustrated in *fig. 8*.

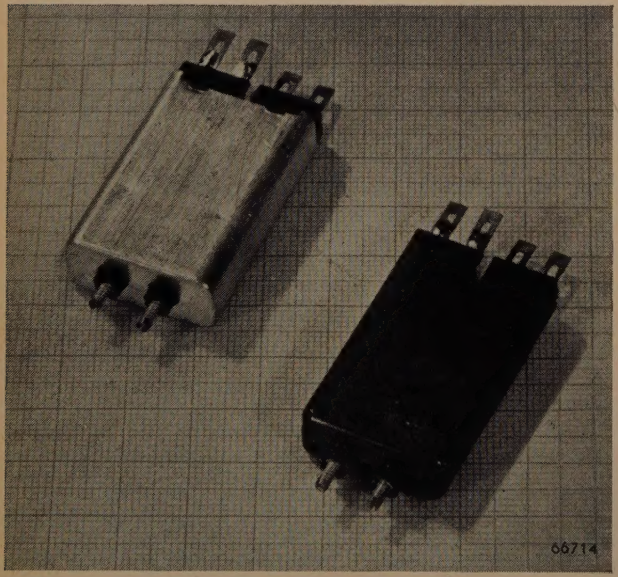


Fig. 8. The type AP 1000 microtransformer. The inductance of the coils can be adjusted with the two screws. The filter on the right has had its screening aluminium jacket removed.

The wire capacitor is likewise often employed in simple high-frequency tuning circuits. These consist of a small coil with a movable ferroxcube core and a wire capacitor. By screwing the core farther in or out the coil it is possible to vary the resonant frequency of the circuit. This device replaces the conventional tuning circuits built up from one or more coils and one or more variable capacitors. *Fig. 9* is an illustration of a tuning unit consisting of a high-frequency circuit and an oscillator circuit.

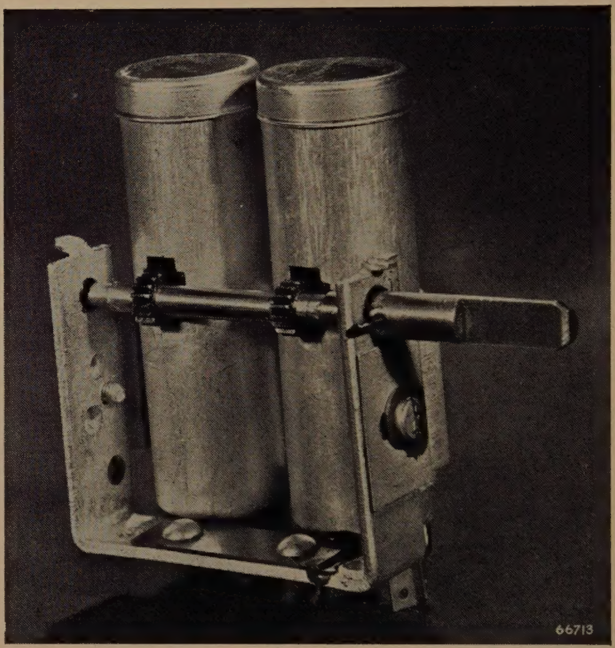


Fig. 9. High-frequency tuning unit, consisting of an aerial circuit and an oscillator circuit. Tuning is done with the aid of the spindle carrying gear wheels, each engaging a toothed rail affixed to the cores of the coils.

Other composite drawn products

As already stated, the drawing process can be employed also for making indirectly heated cathodes. Such cathodes consist of a filament separated by an insulating layer from a metal envelope which in turn is coated with an electron-emitting substance and forms the cathode proper. Thus in principle the construction of the drawn cathode does not differ from that of the wire capacitor, the only difference lying in the materials used and the dimensions.

Owing to the solid construction of the drawn cathode there is very good thermal contact between the filament and the cathode surface. As a consequence the difference in temperature between filament and surface is no more than some tens of degrees, whereas in a normal indirectly heated cathode the temperature of the filament is usually some hundreds of degrees higher than that of the emitting surface. It is therefore possible to use nickel for the filament in a drawn cathode (as also for the outer jacket), so that one is no longer bound to use high-melting metals such as tungsten.

The filament is made as thin as possible so as to minimize the filament current required. The diameter of the outer jacket depends upon the kind of electron valve in which the cathode is to be used; it varies from a few mm to $^1/_4$ mm. Thus it is possible to make a drawn, indirectly heated cathode much smaller than a conventional indirectly heated cathode.

The drawn cathode can be rolled or laminated, forming a ribbon-shaped cathode, without the filament making contact with the jacket. The cathode can also be bent and wound into various shapes. For instance, the cathode can be rolled up into a cylinder, closely resembling a cylindrical, hollow, cathode. It can also be wound into a grid-shaped helix, with the turns spaced apart, which is a great advantage for special applications, or into a flat spiral like a watch spring, closely approximating a flat cathode (see *fig. 10*). Thus the drawn cathode makes it possible for the desired cathode shape to be obtained in a simple manner for all sorts of electron valves,

As other applications of the composite drawing process we would mention here the manufacture of heating elements, of thermo-elements, and the making of vacuum-tight lead-in ports serving, for example, for reaching the inside of an engine cylinder.

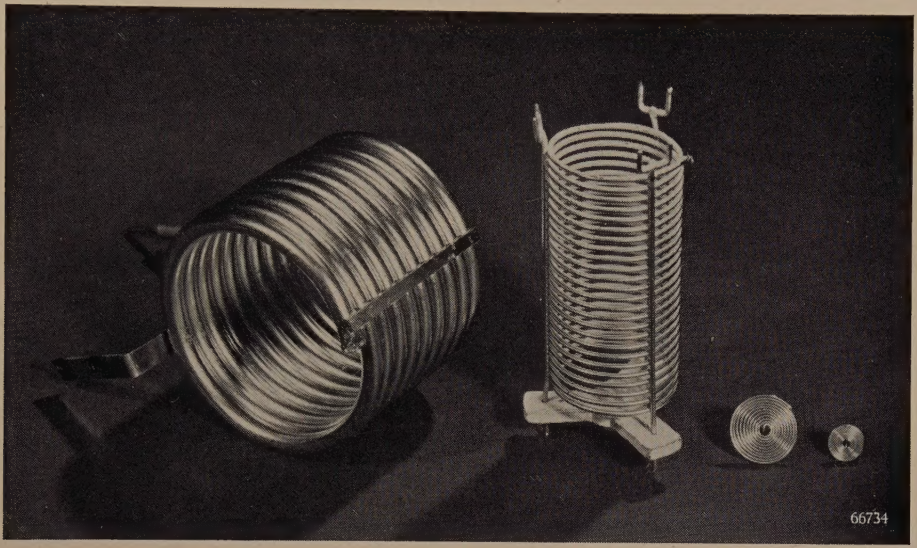


Fig. 10. Various forms into which a drawn cathode can be shaped. From left to right: cylindrical cathode, helical ("grid") cathode, two flat cathodes in the form of a watch spring.

Summary. By means of a special drawing process — the drawing of a tube containing a wire core, with the space in between filled with an insulating material — it has been found possible to manufacture capacitors possessing very good properties. These wire capacitors, being of the smallest dimensions hitherto known, are ideally suitable for use in assembled parts, such as I.F. transformers. The special process of manufacture sets high requirements for the materials from which the capacitor is made. Some details of the manufacture are described, such as the baring of the inner conductor, which can be done either by etching or by shaving on a lathe. A discussion of the temperature-dependency of the capacitance shows that to a certain extent this can be controlled by the choice of material and dimensions. The dielectric losses, which are small, appear to be caused by oxidation of the metal; this makes it necessary to age the capacitors. Some applications of the wire capacitor are discussed, followed by a brief reference to the drawn cathode, an indirectly heated cathode with special properties, e.g. exceptionally easy shaping.

A SIMPLE IONOSPHERE SOUNDER

by R. ASCHEN *) and P. GAILLARD *).

621.396.91

Apart from its great scientific value the investigation of the ionosphere is also of direct practical value, especially for predicting the most favourable wavelength for a radio communication between different points on the earth at specified times. The American Bureau of Standards, for instance, issues monthly predictions, based on investigations of the ionosphere at various places on the earth, from which the most favourable wavelength for a given radio link can be deduced with a fair degreee of certainty. P.T.T. services, radio and airways companies and suchlike are making good use of this information. This explains the increasing interest being shown in the investigation of the ionosphere, a fact which is to be welcomed because the more information that is collected from observations the more accurate will be the predictions.

One of the ionosphere sounders designed by the authors has purposely been kept as simple as possible so that it may be brought within the reach of serious radio amateurs, who — after having "discovered the short waves" — may yet again yield an important contribution towards the common good by regularly carrying out ionosphere soundings.

As may be kown, it is convenient to divide the atmosphere of the earth into zones, the most important of which are called the troposphere, the stratosphere and the ionosphere. The transitional layer, rich in ozone, between the last two zones is called the ozonosphere[1]). *Fig. 1* is a graphical representation of this division. In this article we are concerned only with the ionosphere.

One of the reasons why the investigation of the ionosphere is now being taken up on such a large scale is that the ionosphere has such a considerable influence upon radio traffic. This immediately became a matter of importance at the beginning of this century when Marconi first established transatlantic wireless communication and the question arose as to how the radio waves were able to follow the earth's curvature. The hypothesis — proposed by Kennelly and Heaviside independently of each other — is that in the atmosphere there is an ionized layer acting as reflector for radio waves and making it possible for them to be propagated over long distances in one or more "hops". Though this hypothesis dates from 1902, it was not until 1924 that Appleton succeeded in proving the existence of such a layer, and even of several layers, which he named the D, E and F regions, the last one being split up in the daytime into two layers (F_1 and F_2; see fig. 1).

Appleton carried out his investigations by using radiotechniques, and even at the present day the same method (in a somewhat modified form) is being mainly followed, but it is no longer the only method. Direct exploration by means of instruments sent up into the ionosphere has now been made possible by the adoption in America

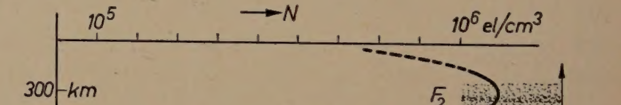


Fig. 1. Division of the atmosphere into the troposphere, the stratosphere, the ozonosphere and the ionosphere. The vertical scale gives the height h. The greatest heights reached by a normal radio sonde balloon (*rs*), a manned balloon (*b*) and — as far as is known — by a V2 rocket are indicated to scale. *D*, *E*, F_1 and F_2 are the ionized regions present in the ionosphere in daytime. At night the D region is entirely absent and the E region has disappeared for the greater part, while the F_1 and F_2 layers are combined into the F region. The height of these layers, especially that of F_1 and F_2, is changeable. The curve gives the electron density N as a function of h.

---

*) Laboratoire d'Applications Electroniques Philips-France, Paris.

[1]) Proposals have been made for a more detailed and more logical classification. See S. Chapman, Upper atmospheric nomenclature, J. geophys. Research 55, 395-399, 1950 (No. 4).

of the technique of V2 rockets. With these rockets it has been possible to measure the concentration of ions at an altitude of 100 km [2]) and also the temperatures in or above the E layer. Unfortunately this is a very expensive technique, even when employing the small rocket ("aerobee") specially designed for physical observations. For this reason alone it is not suitable for use at many places on earth and for a number of times per day, as is necessary if a good picture is to be formed of the state of the whole of the ionosphere and of the changes to which this is continually subject.

For such routine investigations on a large scale it is, therefore, still exclusively radio methods that are to be considered. Five years ago an account of the results of this radio investigation was published in this journal [3]). Some of these results will be reviewed in the present paper, dealing rather more with the practical value and the method of the investigations, and this will be followed by a description of a simple apparatus which has been designed for investigating the ionosphere.

Principle of an ionosphere sounder

For quite a number of years the investigation of the ionosphere by radio has no longer been carried out by the method employed by Appleton (a description of which can be found in the article by Bakker quoted in footnote [3])), but by a system developed by Breit and Tuve, upon which radar is also based. Short wave trains are sent out by a transmitter (*fig. 2a*), and a receiver in the immediate vicinity picks up both the direct wave and the reflections from some object (in the case of ionosphere investigations, the reflections from one of the ionized layers in the atmosphere). While the transmitter is working, the signal received direct has to be limited in order to avoid adverse effects. Both the direct signal and the reflected signal as picked up by the receiver are displayed on the screen of a cathode-ray tube. In the simplest form of the apparatus, which is called the ionosphere sounder, these signals appear as a vertical deflection in the form of pulses, while in the horizontal direction a linear time base is traced, synchronized with the repetition frequency of the wave trains. This frequency is chosen such that when the wave train received direct produces a pulse on the extreme left of the oscillogram, the echo received from the ionosphere produces a pulse on the screen at a measurable distance to the right of the first one fig. 2*b*). The known time-base frequency gives the echo time t_e between the two pulses and from this can be found the altitude h' of the reflecting layer ($h' = \frac{1}{2}ct_e$, where c is the velocity of light). A kilometre scale drawn along the time axis gives direct readings of h'.

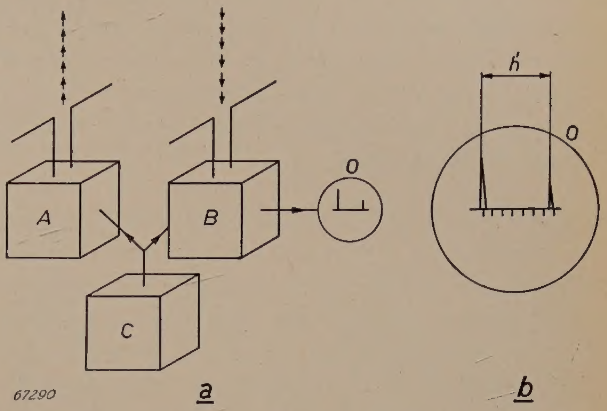


Fig. 2. Schematic representation of the principle of an ionosphere sounder according to Breit and Tuve.
a) *A* the transmitter sending out pulse-shaped wave trains, *B* the adjacent receiver connected to the cathode-ray tube *O*. The pulses for the transmitter and the time-base voltage for the oscilloscope are supplied by the control unit *C*.
b) When the wave is reflected back from the ionosphere two pulses are seen on the oscilloscope, one from the signal received direct and the other as an echo from the ionosphere. The distance between the two pulses is a measure for the (virtual) height h' of the reflecting layer in the case of perpendicular incidence.

Two characteristic phenomena of ionospheric reflection are the following:

1) the quantity h' measured in the manner described appears to be dependent upon the frequency f of the transmitter, and
2) the result varies with the time of day, the season and the phase of the year with respect to the eleven-year cycle of the sun-spots, whilst it also shows irregular fluctuations.

These two effects call for a closer consideration.

Measurement as a function of the frequency

It is not surprising that h' depends upon f when it is borne in mind that the signal is not reflected from a sharply determined surface but from a layer whose electrical properties vary continuously with altitude, this corresponding to a gradually changing refractive index. These electrical properties depend primarily upon the density of the free electrons present in the ionosphere. The influence of the ions, likewise present, upon the propagation of the radio waves can be ignored, because their relatively large mass as compared with that of the electrons makes their effect negligible.

---

[2]) A. Reifman and W. G. Dow, Phys. Rev. **75**, 1311-1312, 1949 and **76**, 987-988, 1949.

[3]) C. J. Bakker, Radio investigation of the ionosphere, Philips techn. Rev. **8**, 111-120, 1946. See also T. W. Bennington, Short-wave radio and the ionosphere, Iliffe & Sons Ltd., London, 1950.

In each of the layers the electron density is greatest and the refractive index smallest in the middle of the layer. An incident ray with frequency f_1, for instance, therefore follows a partly curved path as represented in *fig. 3a*. The greatest height it reaches, $h_{\mathrm{max}1}$, is less than the virtual height h_1' found from the echo time. (It is only for the sake of clarity that in fig. 3 the distance between A and B has been drawn rather large. In practice the transmitter and the receiver for ionosphere sounding are placed so close together that the distance can be ignored, so that the angle of incidence α may be taken as 0.) In the curved part of the path the velocity ⁴) is less than c.

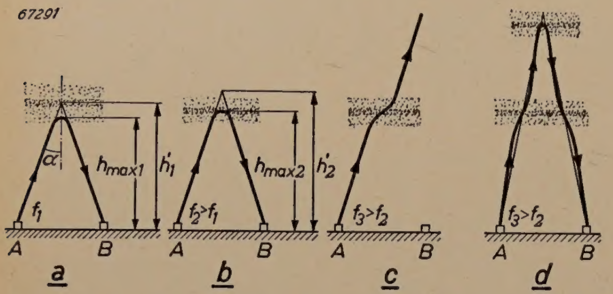


Fig. 3. The path followed by radio waves in ionosphere sounding (here for the sake of clarity the transmitter A and the receiver B are placed so far apart that the angle of incidence α is much greater than it is normally).

In the case (a) the frequency (f_1) is so low that the wave is already turned back to earth when it reaches the bottom of the reflecting layer. In the case (b) the frequency (f_2) is a little higher and the waves penetrate deeper into the layer. In the cases (c) and (d) the critical frequency is exceeded and the waves pass right through the layer. In the case (d) there is a higher layer reflecting the waves back to earth. h_{max} = the greatest height reached by the reflected ray, h' = the virtual height measured.

If a somewhat higher frequency is chosen (f_2) then either a greater density of electrons or a longer refractive path is required to reflect the ray completely. It therefore penetrates deeper into the layer ($h_{\mathrm{max}2} > h_{\mathrm{max}1}$, fig. 3$b$) and a greater virtual height h'_2 is recorded. Upon the frequency being increased still further a limit is reached where the ray penetrates into the zone of maximum electron concentration N and minimum index of refraction. When the frequency exceeds this limit the ray passes right through the layer. The frequency limit in the case of perpendicular incidence ($\alpha = 0$) is called the critical frequency (f_{crit}). From the mechanism of the interaction between the free electrons in the ionosphere and the field of the incident radio wave it is possible to derive a simple relation between the maximum electron density N_{max} and the critical frequency (see, e.g., the article by Bakker quoted in footnote ³)):

$$N_{\mathrm{max}} = 1.24 \cdot 10^{-8} \cdot f^2_{\mathrm{crit}}, \quad \ldots \quad (1)$$

where N_{max} is expressed in cm^{-3} and f_{crit} in c/s.

For a frequency higher than f_{crit} the electron density of the layer is not great enough to deflect the ray back to earth. There are then two possibilities: either the ray disappears into space (fig. 3c) or it is reflected back to earth by a higher layer with sufficient density of electrons (fig. 3d); examples of the latter phenomenon will be given later.

The critical frequency for a certain layer — i.e. the highest frequency at which an echo can be received from that layer at normal incidence — can in most cases be sharply determined. According to eq. (1) the value found for f_{crit} gives at once the greatest electron density in the layer.

From fig. 3 it is seen that the virtual height h' as found by measurement is always greater than the maximum height h_{max} reached by the radio wave. As long as f is much lower than f_{crit} the difference between h' and h_{max} is very small, but at frequencies near f_{crit} the velocity in the layer is much smaller than c, so that the echo time for an observed echo is thereby increased and the value of h' deduced from it is much greater than the height at which the layer has the greatest electron density.

These phenomena find expression in the graph obtained by plotting the measured values of h' as a function of the frequency, which is increased in small steps from, say, 1 to 12 Mc/s. *Fig. 4a* shows the situation as found about noon on a winter day at a temperate latitude.

Between 1 Mc/s and 1.5 Mc/s no echo at all is observed, this being due to absorption of the radio waves in the D region, which particularly develops in daytime. This region is also responsible for the well-known fact that many remotely situated broadcasting stations working on a wavelength between 600 and 200 m cannot be heard in the daytime, whereas in the evening they can be picked up quite well. The D region, having a critical frequency much lower than 1 Mc/s, does not reflect waves with frequencies higher than 1 Mc/s.

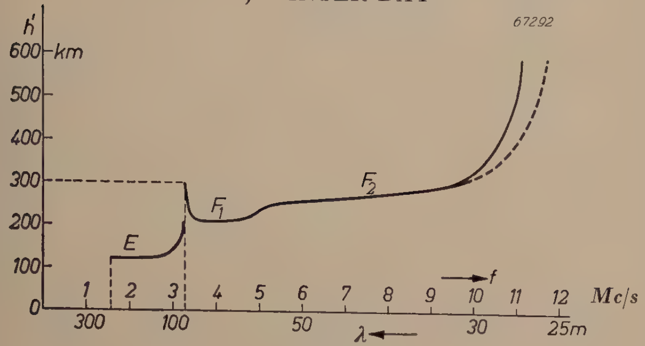
At $f = 1.5$ Mc/s an echo can be observed corresponding to a virtual height of slightly over 100 km: the E layer. This proves that for this and higher frequencies the absorption in the lower-lying D region is sufficiently reduced to make reception possible. As explained with reference to fig. 3, as the frequency is raised so h' gradually increases, until at $f = 3.2$ Mc/s the critical frequency for the

---

⁴) Here the group velocity is meant, and not the phase velocity.

E layer is reached: h' then suddenly increases from about 200 km to 300 km. This means that from this frequency onward the E layer suddenly becomes "transparent" and that there is a layer at a higher altitude — the F_1 layer — reflecting the radio waves

a) *WINTER DAY*



b) *SUMMER DAY*

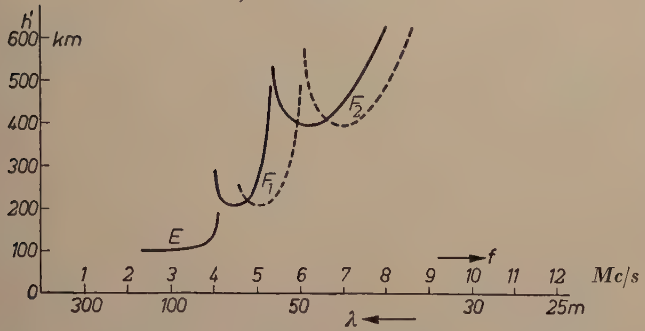


Fig. 4. Examples of sounded virtual heights h' as functions of the frequency f or of the wavelength λ, (*a*) at noon on a winter day, (*b*) at noon on a summer day, both on a temperate geographical latitude. The letters E, F_1 and F_2 denote the layers acting as reflector. The curves drawn in broken lines relate to the "extraordinary wave" formed by double refraction as a result of the earth's magnetism.

(and thus apparently having a greater electron density than the E layer). The great retardation to which signals of about the critical frequency are subjected in the E layer finds expression in the steepness of the curve either side of the discontinuity.

The fact that the greatest value of h' immediately preceding the discontinuity — about 200 km — is much greater than the actual height of the E layer has been proved by other observations, from which it has been possible to conclude that the effective thickness of this layer is no more than a few kilometres (1 to 4 km), whilst the height of its centre is about 120 km.

As the frequency is increased further, h' at first diminishes rapidly (due to reduced retardation in the E layer) but soon begins to rise again gradually. At about 5 Mc/s the critical frequency of the F_1 layer is passed, after which the higher F_2 layer takes over the function of reflector. At this transition there is often no decided discontinuity in the curve owing to the fact that in winter there is no greatly pronounced separation between these two layers. The absence of a long flat part in the curve is due to the fact that the two F layers are much less sharply defined than the E layer.

From about 9 Mc/s onward a peculiar phenomenon arises: the echo pulse in the oscillogram is seen to split up into two parts, corresponding to the two branches of the right-hand part of the curve in fig. 4*a*. This phenomenon is related to the earth's magnetic field causing double refraction, so that an incident wave is reflected back as an "ordinary" and an "extraordinary" wave [5]). It yields two critical frequencies for the F_2 layer, the difference between them (mostly about 0.7 Mc/s) depending upon the terrestrial magnetic field strength at the point of reflection; as far as the present considerations are concerned it may be ignored. The fact that this double refraction is apt to occur also at lower frequencies is proved by the oscillogram in *fig. 5*, recorded at a frequency of 5.75 Mc/s with the ionosphere sounder to be described.

As the frequency is still further increased there is no longer any echo to be observed, from which it is to be concluded that above the F_2 layer there are no other layers having a sufficient density of electrons to reflect the waves penetrating through the F_2 layer.

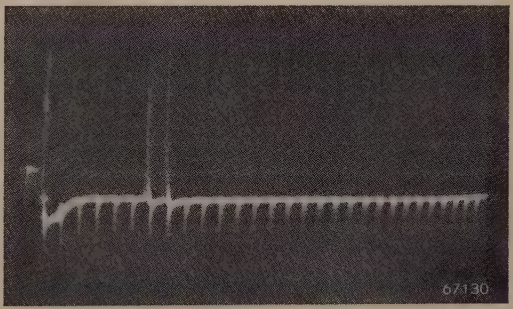


Fig. 5. An oscillogram recorded with an ionosphere sounder (at Paris, 30th March 1951 at 16 hrs) at a frequency of 5.75 Mc/s. On the extreme left the pulse received direct. To the right of that are two echo pulses caused by double refraction in the ionosphere. The interval between the pulses directed downward corresponds to a difference in height of 50 km.

Others factors affecting the results of a sounding

The values given for the critical frequencies in the preceding section are to be regarded as being typical for the situation about noon on a winter day. At other times and in other seasons greatly different values are often found. Considering that in the first place the sun is responsible for the ionization of the atmosphere, it is not surprising that the position of the earth with respect to the sun, and more particularly the local time and the geographical latitude of the point in the atmosphere considered, will be of great influence. Fig. 4*b* gives a graph for $h' = f(f)$ for a temperate latitude at at noon on a summer day. The occurrence of a pronounced critical frequency of the F_1 layer, the great difference in altitude between F_1 and F_2, and the lowered critical frequency of F_2 are factors which contrast with the situation in winter.

---

[5]) See, i.a., the literature cited in footnote [3]).

Fig. 6 gives an idea of the variation in the course of 24 hours for a summer day and a winter day. The expectation that the ionization will be most intense at the highest position of the sun is confirmed with respect to the E and the F_1 layers. The fact that, contrary to expectations, in the course of a summer day the critical frequency of the F_2 layer does not reach such a high maximum as on a winter day may be related to the rising (and thus expansion) of this layer on a summer day, the resultant reduction in electron density being greater than the increase caused by ionization by the sun's rays.

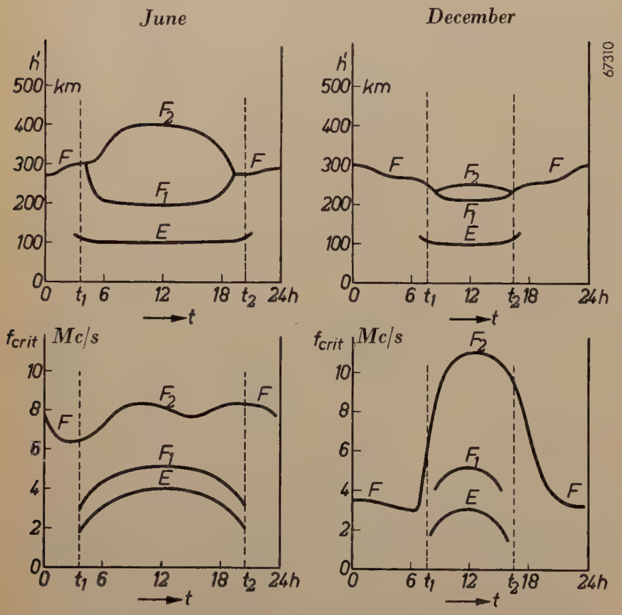


Fig. 6. Some results of periodical ionosphere soundings. Changes taking place in the virtual height h' (upper graphs) and in the critical frequencies f_{crit} (lower graphs) of the E and F layers during 24 hours. h' is the virtual height in the roughly horizontal part of the characteristic $h' = f(f)$, thus at $f < f_{\text{crit}}$. The graphs on the left were recorded in June, those on the right in December, both in a year when there was a large number of sun-spots, and at a northern temperate latitude. t_1 = the time of sunrise, t_2 = sunsset.

Conditions in the ionosphere are also closely connected with the number of sun-spots, which, as is known, follows an average cycle of 11 years. Fig. 6 applies for a period when there are many sun-spots. When there are only a few sun-spots as a rule the critical frequencies are lower (especially that of the F_2 layer).

Importance of ionosphere soundings for radio communications

The direct rays of a transmitter have only a very limited range. Radio communication over any great distance can be established only with the aid of the ionosphere ⁶). Here we shall confine our considerations to what are known as one-hop transmissions, where the wave is reflected back to earth by the ionosphere only once. The maximum distance D that can be covered in one hop depends upon the height H of the reflecting layer (for the sake of simplicity the thickness of the layer will be ignored) and the radius R of the earth. From an elementary calculation it follows that $D \approx 2\sqrt{2HR} = 226\sqrt{H}$ (D and H expressed in kilometres), from which it is found that $D \approx 2500$ km for $H = 120$ km (E layer) and almost 4000 km for $H = 300$ km (F layer).

This formula represents the situation rather too favourably, since it applies for waves leaving the transmitter in an absolutely horizontal direction (angle of elevation zero). Such waves, however, are strongly absorbed by the earth. In practice, therefore, an angle of elevation of at least a few degrees has to be reckoned with, and D is then less than $226\sqrt{H}$.

What counts only is the state of the ionosphere at the point of reflection, which is above the point lying half way between the transmitter P and the receiver Q (see *fig. 7*). If we set up at that half-way point an ionosphere sounder A-B and with it determine one of the critical frequencies — for instance that of the F layer — then that f_{crit} is by no means the maximum frequency with which communication can be established from P to Q (at a particular time and via the F layer). The reason for this lies in the difference in the angle of incidence a, which with the ionosphere sounder is zero but much greater for the connection P-Q. For the maximum usable frequencies (M.U.F.) $f_{\max}$, (again ignoring the thickness of the layer):

$$f_{\max} = f_{\text{crit}}/\cos a \quad . \quad . \quad . \quad . \quad . \quad (2)$$

The angle a depends upon the distance P-Q and the height H of the layer in question. Thus it is seen that eq. (2) contains the two data that the ionosphere sounder gives for a certain layer (f_{crit}

---

⁶) Communications by means of centrimetric or decimetric waves via a tropospheric "radio duct" are left out of consideration here.

and h', here $= H$), these together determining the maximum usable frequency with which a given distance (provided it is less than D) can be covered in one hop. Actually, however, owing to the considerable depth to which the waves penetrate into the more or less diffuse layer, the dependency of the relation between $f_{\max}$ and f_{crit} — the "M.U.F. factor" — upon the angle α is much more complicated than that given by eq. (2). The value of this M.U.F. factor lies between 1 and about 5, so that under certain conditions it is possible to work with frequencies much higher than f_{crit}.

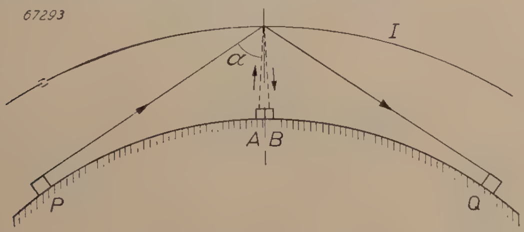


Fig. 7. A one-hop radio transmission from a transmitter P to a receiver Q is dependent upon the state of the ionosphere I halfway between P and Q and upon the angle of incidence α. A-B is an ionosphere sounder set up halfway between P and Q.

What, then, is the most effective frequency for a certain transmission over a distance of some thousands of kilometres? It has to be taken into account that over such a long distance part of it will nearly always lie in daylight, so that, owing to the absorption in the D region of waves shorter than 200 m greatly diminishing with increasing frequency, the frequency has to be chosen as high as possible. For radio-telegraphy, radio-telephony and broadcasting over great distances 6 Mc/s (wavelength 50 m) is considered to be about the lowest frequency. Preferably the frequency is chosen fairly close to the maximum usable frequency $f_{\max}$. According to an American calculation — described below — the optimum working frequency in respect to the F_2 layer is taken to be 0.85 $f_{\max}$ and that for the more stable E layer 0.97 $f_{\max}$ (here $f_{\max}$ is a predicted value).

As we have seen, for the differently ionized layers there are different values of f_{crit}, and for communication between given points there are therefore, according to eq. (2), also different values of $f_{\max}$. Here it is a question of the greatest $f_{\max}$. As a rule this will be the $f_{\max}$ for the F_2 layer (at night the F region).

Sometimes, however, the E layer may be decisive. At noon on a summer day, for instance, f_{crit} of the E layer is only slightly lower than f_{crit} of the F_1 layer (see fig. 6). For a one-hop transmission between given points however the angle of incidence on the E layer is greater than that on the F_1 layer ($\alpha_E > \alpha_F$, *fig. 8*), thus also $\cos^{-1}\alpha_E > \cos^{-1}\alpha_F$, so that it is not precluded that the product of f_{crit} and $\cos^{-1}\alpha$ for the E layer is greater than that for the F_1 layer.

In the last ten years the Americans have taken the lead in working out a system for predicting $f_{\max}$ values from periodical ionosphere soundings taken at a large number of places on the earth. The Central Radio Propagation Laboratory of the National Bureau of Standards at Washington issues graphs with a large number of "predicted" curves along each of which $f_{\max}$ has a constant value; the coordinates are the geographical latitude and the local time at the point of reflection. Separate sheets with $f_{\max}$ contours are issued for a certain ionospheric layer, for a certain distance and for a particular terrestrial zone (the latter with a view to taking into account, roughly, the influence of the terrestrial magnetism). The curves are derived by extrapolation of the mean variation of values of f_{crit} obtained from numerous ionosphere soundings. They are valid for one month and are published three months in advance.

As regards the distance, the data given in these graphs are limited to the distances zero (at which $f_{\max} = f_{\text{crit}}$), 2000 km (only for the E and E_s layers[7]) and 4000 km (only for the F region and F_2 layer). With the aid of (rather complicated) directions for use, a world map and nomograms, the optimum working frequency for any transmission can be predicted with a fair degree of certainty.

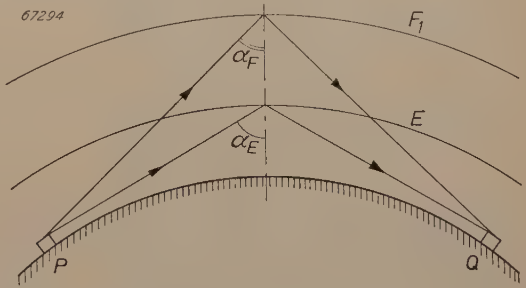


Fig. 8. In the case of a one-hop radio transmission from P to Q via the E layer the angle of incidence is greater than that for a similar transmission via the F region ($\alpha_E > \alpha_F$). Consequently under certain conditions a higher frequency can be used for the transmission via the E layer than for that via the F region.

The foregoing should be sufficient to show the scientific and practical values of ionosphere soundings. We shall now proceed to describe one of the sounders that we have developed. This has been designed mainly with a view to making it as simple as possible, so that it may be within the reach of advanced amateurs, who may furnish a welcome contribution to the information being collected from observations. It is not the intention to go deeply into technical details here; any readers interested in such details are referred to an article published elsewhere [8]), whilst data can be had by applying to the laboratory mentioned in the footnote *) on p. 152.

---

[7] E_s denotes the sporadic E layer, which occurs more or less locally at irregular times. Its height is roughly equal to that of the normal E layer but it has a much greater electron density (thus also a greater f_{crit}).

Description of a simple ionosphere sounder

Fig. 9 is a photograph showing this simple ionosphere sounder. It was with this apparatus that the oscillogram in fig. 5 was obtained. The frequency is variable from 5 to 10 Mc/s. This band is quite sufficient for studying the main properties of the F layers (cf. fig. 6) and it enables us to observe echoes from the sporadic E_s layer (cf. footnote [7])) when this layer is present. For the repetition frequency of the wave trains the mains frequency has been chosen (50 c/s); the ionospheric conditions do not change so quickly as to justify any higher frequency.

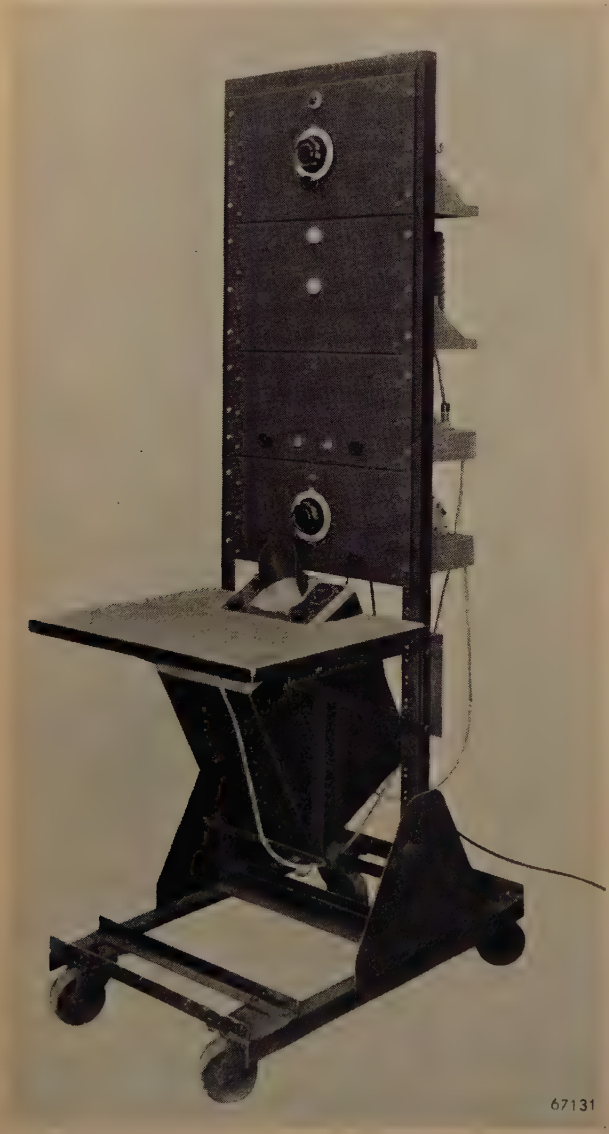


Fig. 9. An ionosphere sounder of simple construction. From top to bottom; the transmitter, the modulator, the control unit, the receiver and type MS 476 oscilloscope.

The wave trains must not be so short as to make it necessary for the circuits to be given an exceptionally large band width. On the other hand the wave train must be shorter than the echo time t_e that the signal takes to travel the distance from the earth to the lowermost layer to be sounded and back to earth again; if this condition were not satisfied then the echo pulse on the screen of the oscilloscope would interfere with the pulse received direct. Since the lowermost layer with which we are concerned is the E_s layer, at an altitude of about 100 km, the time t_e is at least about 700 μsec. For the duration of the wave trains a much shorter time has been chosen, in the order of 100 μsec, thus ensuring a clear separation of the two pulses without it being necessary for the transmitted band to be exceptionally wide.

The principle of the apparatus is entirely in accordance with the explanation given with reference to fig. 2, the oscilloscope showing a pulse received direct and an echo pulse. Consequently the main components to be distinguished are the control unit, the transmitting unit, the receiving unit and the aerials.

Control unit

The control unit has to supply the signals for coordinating the working of the transmitter with that of the receiver. Every 1/50th second it produces a pulse (see *fig. 10*) which brings the transmitter into action, a linearly increasing and decreasing voltage for the time base of the cathode-ray tube, a voltage for blanking out its beam during the flyback, and a series of pulses forming a distance scale along the horizontal axis.

The wave forms and phase relations of the various signals are represented in fig. 10. All these signals are derived from one single square-wave voltage, called the control voltage, so that they are all perfectly constant in phase (the phase of the pulses controlling the transmitter is slightly variable; this will be dealt with presently).

The following circuits are to be distinguished:
a) Control circuit. The square-wave voltage controlling the whole apparatus is derived from a sinusoidal alternating voltage of 50 c/s. This is taken, via a resonant circuit tuned to 50 c/s (see *fig. 11*), from a winding of the supply transformer and applied, via a resistor of 0.5 MΩ, to the grid of the first triode section of an ECC 40 valve connected in cascade with the second section. In this

---

[8]) P. Maguer, Le sondage ionosphérique, Radio Française 1949 (No. 4), pp. 7-12.

circuit a large part of the two sine peaks is clipped. At the anode of the second triode section a practically rectangular voltage is obtained, the wave form of which shows steep flanks; this voltage is suitable for use as control voltage.

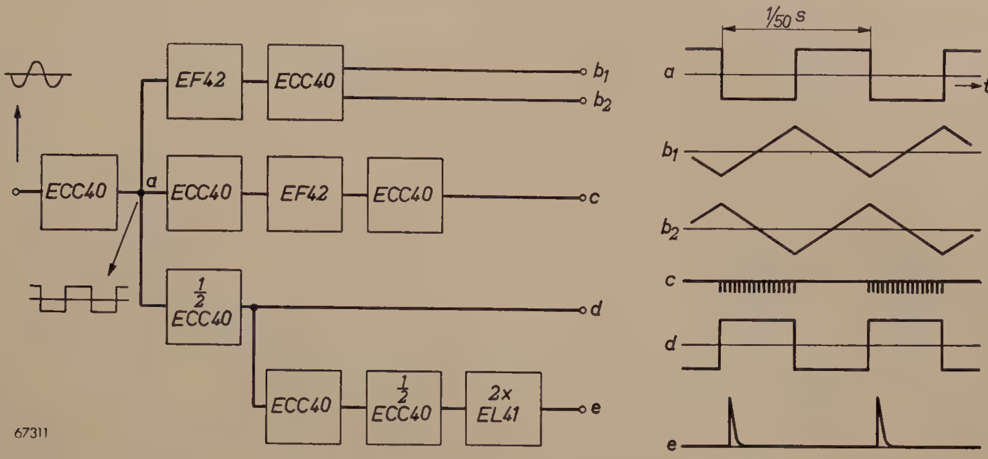


Fig. 10. Block diagram of the control unit. The ECC 40 valve (on the extreme left), to which is applied a sinusoidal voltage of 50 c/s, produces a square-wave voltage (control voltage) as represented at (*a*). From this control voltage are derived: the push-pull voltage for the horizontal deflection (b_1, b_2), the voltage for the electronic altitude scale (*c*), the voltage for periodically blanking out the cathode ray (*d*), and the pulses for periodically starting the transmitter (*e*).

The accuracy of the synchronizations brought about by the control voltage is proportional to the steepness of the flanks. The above-mentioned resonant circuit removes any harmonics, which might distort these flanks.

b) **Circuit for generating a triangular deflecting voltage.** This circuit (*fig. 12*) supplies a voltage for the horizontal deflection in the cathode-ray tube. It comprises, i.a., an integrator fed with the square-wave control voltage, so that the output voltage assumes the shape of an isosceles triangle (fig. 10, b_1).

The integrator is of the Miller type [9]). It contains an EF 42 pentode, the control grid of which is fed via a resistor R_2 of 2 megohms and is connected to the anode via a capacitor C_2 of 50,000 pF. The time constant R_2C_2 is 0.1 sec, thus greatly exceeding the duration of a cycle of the control voltage to be integrated. With this value of R_2C_2 the form of the output voltage deviates less than 1% from linearity.

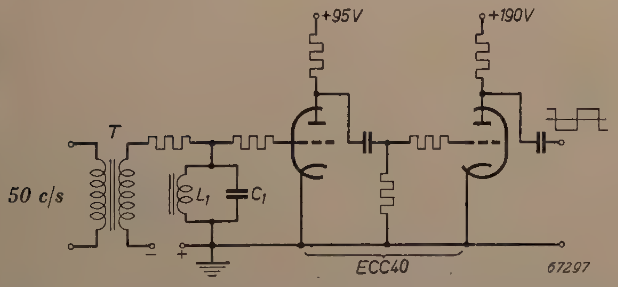


Fig. 11. Circuit diagram of the control unit. *T* = mains transformer. L_1-C_1 = circuit tuned to 50 c/s. ECC 40 a double triode in which both peaks of the sinusoidal voltage are clipped off, producing a square-wave voltage (the control voltage).

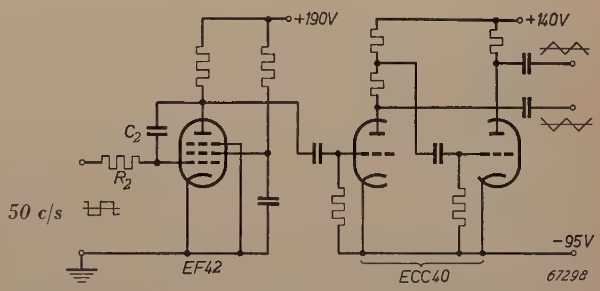


Fig. 12. Circuit for generating the deflection voltage. The pentode EF 42, the resistor R_2 and the capacitor C_2 form a Miller integrator. With a square-wave voltage applied to the input this supplies a voltage having the shape of an isosceles triangle. This voltage is amplified by an ECC 40 valve with balanced output.

The integrator is followed by an amplifier, fitted with an ECC 40 valve. One triode section of this valve amplifies a proportion of the output voltage from the other section, in such a way that the two output voltages are symmetrical with respect to earth (see fig. 10, b_1 and b_2).

c) **Electronic altitude scale.** The electronic graduation is formed by pulses repeated at a frequency of 3000 c/s, the distance between two pulses corresponding to a difference in height of 50 km.

These pulses are obtained with the aid of an oscillator with "phase restoration". This consists mainly of a circuit L_2-C_3 tuned to 3000 c/s in connection with the two triode sections of an ECC 40 valve (*fig. 13*). Together with this tuned circuit the right-hand triode section (*I* in fig. 13) forms a Hartley oscillator. A variable resistor R_3, connected between the cathode and the tapping on L_2, is so adjusted that the oscillations are just maintained.

---

[9]) The Miller integrator has been discussed, i.a., in Philips techn. Rev. **12**, 328-330, 1951 (No. 11), fig. 15 and footnote [7]).

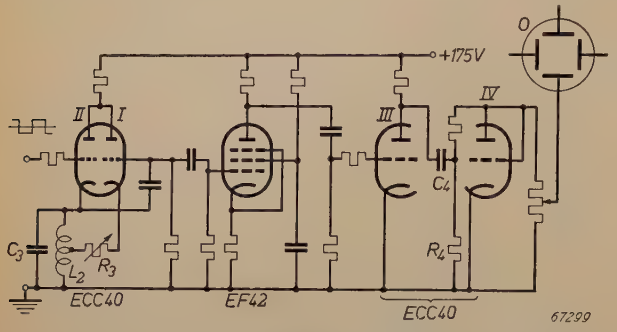


Fig. 13. Circuit for the electronic altitude scale. The circuit L_2-C_3, tuned to 3000 c/s or 1500 c/s, and the triode section *I* of an ECC 40 valve form a Hartley oscillator. Its phase is restored 50 times per second by the section *II* of the same valve, to the input of which the control voltage is applied. EF 42 is an amplifying valve, *III* a triode clipping the peaks. C_4-R_4 a differentiator, *IV* a diode suppressing the pulses of undesired polarity. *O* cathode-ray tube.

Generally the frequency of these oscillations is not exactly a multiple of 50 c/s, so that some device is required to ensure that the oscillator always commences to oscillate with the same phase, with respect to the flanks of the control voltage, each time the horizontal axis begins to be traced. This restoration of the correct phase is brought about by the other triode section (*II*) of the ECC 40, to the grid of which the control voltage is applied, in such a way that the oscillator is started and stopped 50 times per second. At the moment that the control voltage changes from positive to negative, the anode current of *II*, flowing through the coil L_2, is suddenly interrupted and the circuit L_2-C_3 begins to oscillate. With the aid of R_3 the damping is reduced almost to zero, so that the amplitude of the oscillations is practically constant (the transient is of very short duration) up to the end of the half-cycle in which the control voltage is negative. Then the triode *II* is again unblocked, and since its small internal resistance is shunted across the oscillatory circuit the oscillations are quickly damped. The oscillations begin anew, in the resonant frequency of the circuit L_2-C_3, as soon as the control voltage again changes from positive to negative.

The sinusoidal voltage of 3000 c/s is taken from the grid of the triode section *I*. It is amplified by an EF 42 pentode and the peaks of the sine are clipped in the usual way by the triode ECC 40 (*III*), the output from which is thus a square-wave voltage of 3000 c/s, This is applied to a circuit formed by a capacitor C_4 and a resistor R_4 connected in series. The time constant of this circuit is small compared with $^1/_{3000}$ sec. Thus the circuit functions as a differentiator, so that the voltage across the resistor has the form of the derivative of the square-wave voltage applied, viz. a series of pulses of alternating polarity. The second triode section of the ECC 40, connected as a diode, cuts off the negative pulses, so that only positive pulses remain, at intervals of $^1/_{3000}$ sec. When these pulses are applied to the lowermost deflection plate of the cathode-ray tube they produce along the horizontal axis graduation marks, directed downward, corresponding to differences in height of 50 km. (The reason for cutting off the negative pulses is that these would produce graduation marks directed upward, which might easily be confused with the likewise upward-directed echo pulses).

The oscillator can be made to work also at a frequency of 1500 c/s, in which case a scale is obtained with graduations of 100 km.

d) Blanking out the beam during the flyback. The control voltage is also applied to one half *V* of another ECC 40 valve (*fig. 14*) so that it supplies a square-wave voltage of opposite polarity. This is applied to the control grid of the cathode-ray tube and blanks out the beam during the half-cycles when it is negative (thus when the control voltage is positive). This makes it possible for the oscillogram of what takes place in the other half-cycle to be spread over the whole width of the screen.

e) Circuit for the intermittent working of the transmitter. This circuit supplies synchronization pulses to a multivibrator which modulates the high-frequency oscillator. These pulses, too, are derived from the control voltage. Unless certain measures were taken, however, owing to the differences in time constants of the various parts of the apparatus, these pulses would not exactly coincide with a pulse of the 3000 c/s oscillator. The transmitting pulse, therefore, would not exactly coincide with a division mark of the altitude scale, and this would affect the reading. To provide against this the pulse which starts the transmitter is passed through with a small variable delay after the control voltage has passed through zero.

The circuit employed for starting the transmitter consists of a flip-flop circuit, providing for the variable delay, an amplifying stage and a pulse generator.

The flip-flop circuit contains an ECC 40 valve the two halves of which (*VII* and *VIII*) form a blocked multivibrator. In the state of rest *VII* is blocked while *VIII* is conducting. From the square-wave voltage at the point *b* — the same voltage that controls the cathode-ray tube — the differentiating circuit C_5-R_5 forms a series of pulses, of which the positive ones cause the triode *VII* to conduct, while the triode *VIII* is blocked. After a certain interval of time the system returns to the state of rest, and his interval can be varied by means of the resistor *P* (fig. 14), across which the capacitor, which couples *VII* to *VIII*, is discharged.

The output from *VIII* is amplified by a triode *VI* (forming part of the same ECC 40 valve of which the part *V* has already been mentioned) and differentiated by a circuit C_6-R_6. At the instants that the flip-flop circuit returns to the state of rest the voltage across C_6-R_6 has a steep front and by differentiation this yields a positive pulse at the point *g*. This is applied to the control grids of two EL 41 valves connected in parallel. Normally these valves are blocked, but when they are triggered by the pulses they cause a capacitor C_7 of 0.5 μF to discharge through a winding of a transformer forming part of the modulator (see below).

f) Supply. The D.C. supply for the control part is provided by a rectifier equipped with an 1882 rectifying valve. The direct voltage produced is smoothed and then stabilized by means of three 4687 neon tubes connected in series. The point between the first and the second stabilizers is earthed. This yields a positive voltage of about 190 V and a negative one of 95 V serving as grid bias. The supplies for the various parts are by passed to earth.

Transmitting unit

The transmitting unit (*fig. 15*) comprises the modulator, the supply unit and the transmitter proper. The latter consists of a self-oscillator fed with voltage pulses.

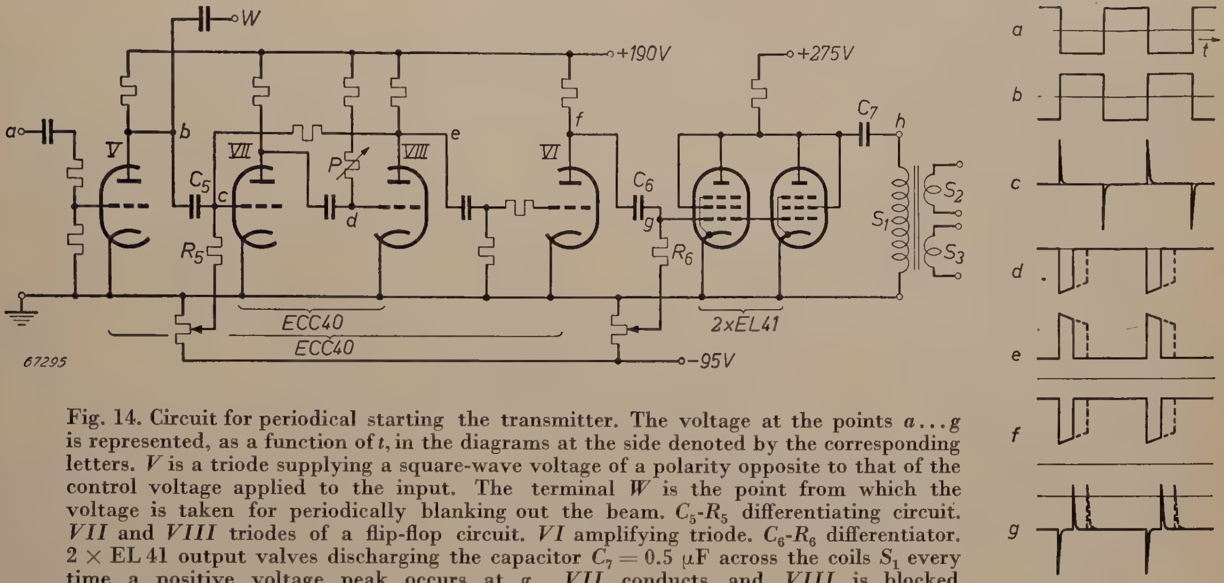


Fig. 14. Circuit for periodical starting the transmitter. The voltage at the points *a* ... *g* is represented, as a function of *t*, in the diagrams at the side denoted by the corresponding letters. *V* is a triode supplying a square-wave voltage of a polarity opposite to that of the control voltage applied to the input. The terminal *W* is the point from which the voltage is taken for periodically blanking out the beam. C_5-R_5 differentiating circuit. *VII* and *VIII* triodes of a flip-flop circuit. *VI* amplifying triode. C_6-R_6 differentiator. 2 × EL 41 output valves discharging the capacitor $C_7 = 0.5$ μF across the coils S_1 every time a positive voltage peak occurs at *g*. *VII* conducts and *VIII* is blocked each time a positive peak occurs at *c*. The original position (*VIII* conducting, *VII* blocked) is restored a certain time later, this time being adjustable by means of the variable resistor *P*. Thus the pulses at *d*, *e*, and *f* are of variable duration, and the positive peaks at *g* have a variable phase.

The anode of the modulator valve (TB 2/200 [10]) is fed, via the coil S_2 of a transformer with three windings, with 1100 V direct voltage supplied by a rectifier equipped with a DCG 4/1000 mercury-vapour rectifying valve. Current pulses from the control unit are fed to the coil S_1 (fig. 14). In the grid circuit the coil S_3 and a negative bias source are included.

Each pulse produces an anode current pulse in the otherwise blocked modulator valve [11]). The coupling between S_2 and S_3 is such that each time this valve is triggered an oscillation is produced which lasts only one cycle. The sudden interruption of the anode current produces on the anode of the valve a large voltage pulse of about two thousand volts which causes the transmitting oscillator to work momentarily; during the rest of the time this oscillator remains inoperative.

---

[10] The TB 2/200 valve of the modulator can to advantage be replaced by a QB 3/300 tetrode and the two TB 1/60 valves of the oscillator by a QQE 06/40 double tetrode (for a description of the latter see: E. G. Dorgelo and P. Zijlstra, Two transmitting valves for use in mobile installations, Philips techn. Rev. **12**, 157-165, 1950 (No. 6)). These more modern types were not available at the time that this ionosphere sounder was designed.

In the apparatus described here the maximum peak values of the anode current and anode voltage permissible for continuous operation are greatly exceeded, so that the normal guarantee for the valves does not hold in this case.

[11] For a more detailed description of this circuit see, e.g. L. Liot, Le générateur d'impulsions à auto-excitation, Electronique, April 1951 (No. 53), pp 8-11.

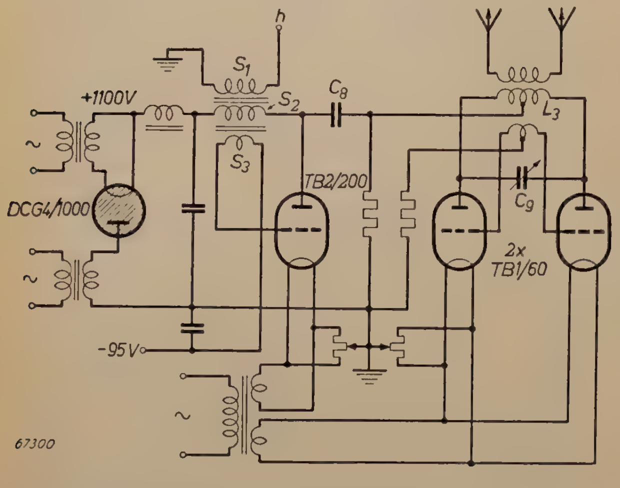


Fig. 15. Circuit diagram of the transmitter. TB 2/200 modulating valve discharging the capacitor C_8 = 0.25 μF via the Mesny oscillator (valves TB 1/60 and oscillator circuit L_3-C_9). each time a current pulse is sent through the coil S_1 (cf fig. 14). By means of C_9 the frequency can be adjusted between 5 and 10 Mc/s.

Receiving unit

The receiver is a sensitive superheterodyne receiver (see *fig. 16*) containing an H.F. amplifying stage with a symmetrical input and equipped with an EF 42 valve. As mixer a triode-hexode ECH 41 is employed. The I.F. amplifier is adjusted to 472 kc/s and comprises two stages each with an EF 42 valve. The bandwidth of the I.F. transformers is larger than that of an ordinary receiver, in order to obtain a better pulse response. The band covers 18 kc/s between the frequencies at which the gain drops 6 db.

Detection takes place in a diode EB 40 in combination with an adjustable threshold voltage. The latter enables interferences to be eliminated when the field strength of the ionospheric echo is sufficiently strong, The A.F. amplifier has an EF 41 valve loaded with a resistor. The output voltage provides for the vertical deflection of a type MS 476 cathode-ray oscilloscope.

In this apparatus there was no need to take steps for purposely blocking the receiver for the direct rays while the transmitter is working. With the low anode voltages in the receiver (60 V) the signal received direct is already sufficiently limited to avoid any adverse effects.

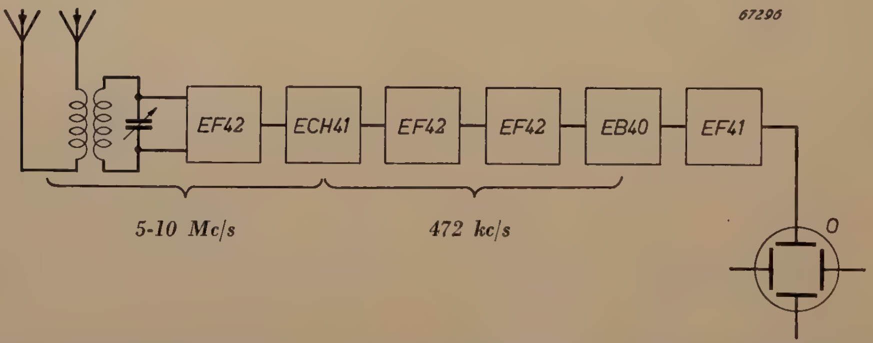


Fig. 16. Block diagram of the receiver, consisting of an H.F. amplifying stage (EF 42), a mixing stage (ECH 41), two I.F. amplifying stages (EF 42), a detector (EB 40) and an A.F. amplifying stage (EF 41) connected to the oscilloscope *O*.

The transmitting oscillator works with two TB 1/60 [10]) triodes in a symmetrical circuit with inductive feedback (Mesny circuit). The peak output is about 2 kW. As already mentioned, the transmitting frequency is variable from 5 to 10 Mc/s. The tuned circuit is formed by a coil wound on a ceramic former (likewise carrying the grid coil) and a variable capacitor of 130 pF insulated for 5000 V (the H.F. voltage reaches a peak value of 4000 V).

Aerials

The transmitting and receiving aerials, which are exactly alike, have been so designed as to give a vertical directional effect and to require only two points of suspension.

Each aerial (see *fig. 17*) has the shape of a rectangle 1 m high and 40 m long in the horizontal direction (i.e. $^2/_3$ to $^4/_3$ times the wavelength at which the apparatus works). The two aerials are spaced 1 m apart. In the middle of the wire along the bottom of each rectangle is an insulator, on either side of which a feeder 10 m long is connected. At the other end these feeders are connected to two coils, one coupled to the transmitter and the other to the receiver. The impedance of the feeders is about 1000 ohms.

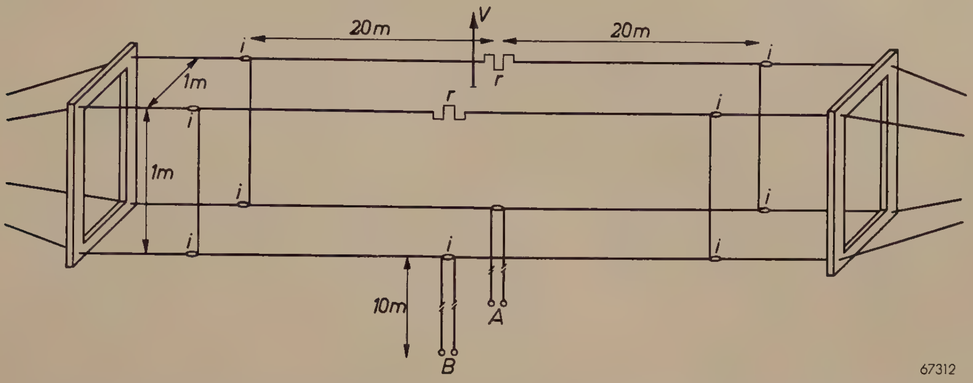


Fig. 17. Transmitting and receiving aerials, each 40 m long and connected via a 10-m lead respectively to the transmitter *A* and the receiver *B* · *i* = insulators, *r* = resistors of 800 ohms (2 W). The arrow *V* denotes the direction of the vertical.

In the middle of the upper wires of the rectangles is a resistor of 800 ohms. These resistors, in which a power of a few watts is dissipated, suppress stationary waves. In this way a current distribution is obtained which results in a maximum radiation to (and reception from) the zenith.

The form of aerial described is a special adaptation of the more common rhombic aerial *(fig. 18.)*. Calculations show that the maximum directional effect with respect to the zenith is obtained when the top half-angle β is 90°. Thus we arrived at the shape of aerial illustrated in fig. 17.

In addition to the apparatus described here, another has been developed in which various improvements and refinements have been incorporated, so that, for instance, the diagram $h' = f(f)$ is produced direct on the screen of the oscilloscope. We hope to be able to describe this type of sounder on another occasion.

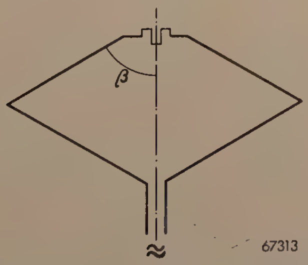
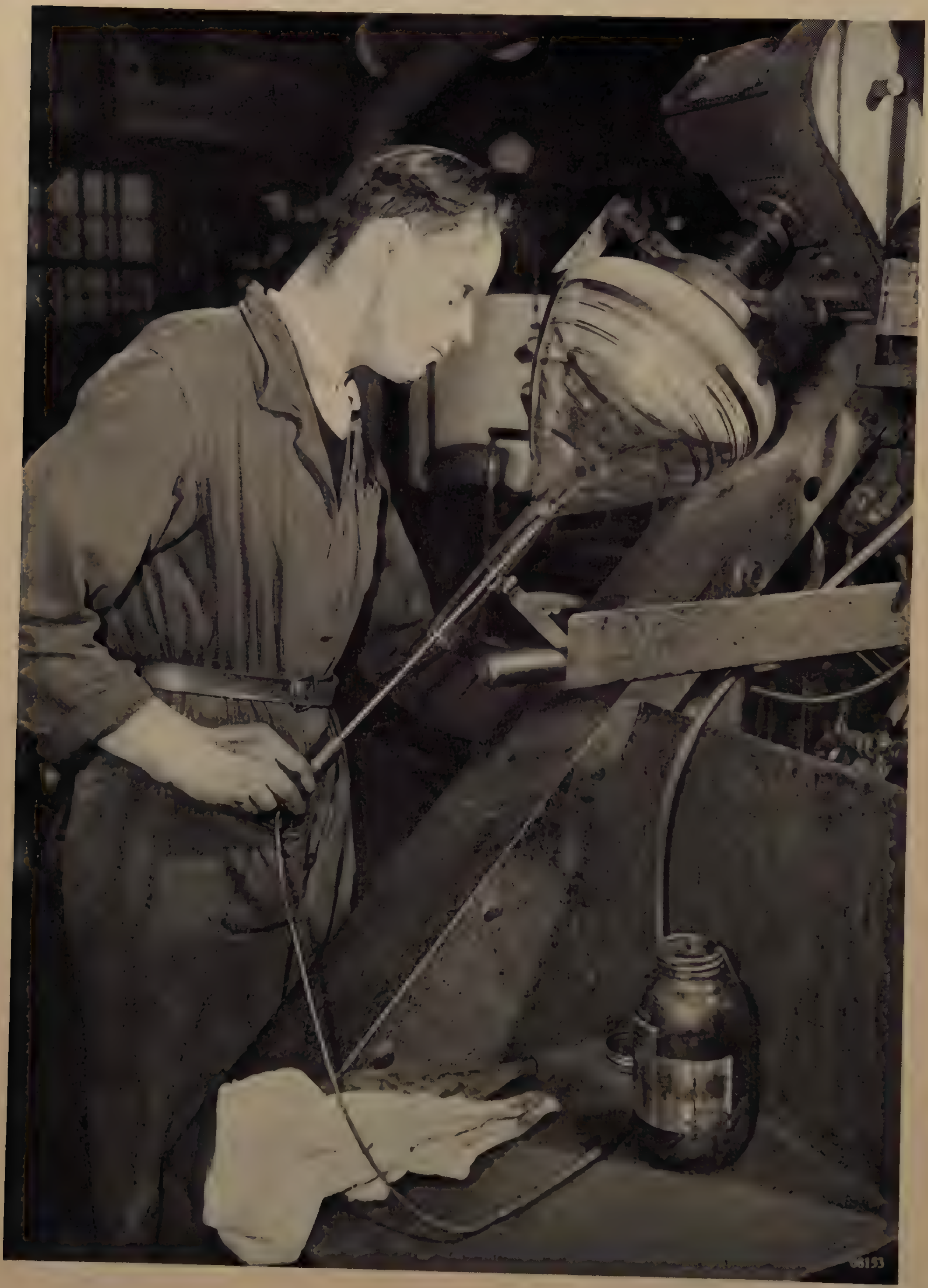


Fig. 18. Rhombic aerial. By making the top half-angle $\beta = 90°$ an aerial is obtained of the shape illustrated in fig. 17.

Summary. Investigations of the ionosphere are mainly carried out by means of the ionosphere sounder according to Breit and Tuve. By this method the echo time elapsing between the transmission of a wave train and the reception of the signal reflected back by the ionosphere in a receiver set up very close to the transmitter is measured by means of an oscilloscope. The measured echo time gives the virtual height of the reflecting layer, and from the highest frequency at which the layer in question still reflects the signal it is possible to find the maximum electron density in that layer. These two quantities as functions of time show various cyclic changes and also irregular fluctuations, whilst they are also dependent upon the geographical latitude of the point of observation. Once the curves for these quantities at a number of places on the earth are known it is possible to predict, some months in advance and with a fair degree of accuracy, the optimum working frequency for any radio transmission. The more data that can be collected from observations of ionospherical conditions, the more reliable will such predictions become. With this end in view the authors of the present article have developed an ionosphere sounder of simple construction, intended for the investigation of the F layers and the sporadic E layer.

The transmitter works with two TB 1/60 valves in a Mesny circuit at a frequency variable between 5 and 10 Mc/s and produces 50 wave trains per second with a duration of about 100 μs and a peak power output of about 2 kW. The receiver is of the superheterodyne type and has a bandwidth of 18 kc/s. The signal received direct and the echo signal are displayed on the screen of an oscilloscope as two pulses. The distance between the two pulses is a measure for the virtual height of the reflecting layer, and this height can be read direct from an "electronic altitude scale". The working of the transmitter is synchronized with that of the receiver by means of control signals all derived from the mains voltage. The aerials give maximum radiation to (and maximum reception from) the zenith.

FROM THE MANUFACTURE OF CATHODE-RAY TUBES FOR TELEVISION RECEIVERS



The glass bulb of cathode-ray tubes has to be coated on the inside with a conducting layer for collecting the secondary electrons emitted by the fluorescent screen; this layer is connected to the anode of the electron gun.

The photo shows a picture tube being coated with such a layer. Aquadag — a suspension of colloidal graphite in a solution of certain organic substances in water — is fed through a rubber tube to a small brush affixed to a bent leaf-spring and passed through the neck of the bulb. The bulb is centred by three guides and rotated about its axis by the action of the atmosphere pressing it against a rotating rubber ring, the space between the ring and the bulb being kept below atmospheric pressure. Starting from the top, the brush is gradually withdrawn, so that first the conical part of the bulb and then the neck is evenly coated with aquadag. The operator regulates the supply of aquadag with his left hand.

AGEING PHENOMENA IN IRON AND STEEL AFTER RAPID COOLING

by J. D. FAST. 620.193.918:669.14-156:621.785.618:621.791.75

After mechanical working or thermal treatment commercial steels often show a peculiar behaviour. For instance, after rapid cooling (quenching), in course of time a steel often shows phenomena of ageing, i.e. changes in its mechanical properties such as its hardness. These phenomena are apt to occur, inter alia, after arc welding, and they appear to depend upon constituents, such as carbon and nitrogen, present in the metal. The atomic processes responsible for this ageing will be discussed in two articles. The first of these — the present one — deals with quench ageing as determined by hardness measurements. In the second article, following upon the first, the processes causing this phenomenon are discussed from an entirely different point of view. A third article in this series, to be published later, will deal with the ageing taking place after mechanical deformations (strain ageing).

Introduction

The normal commercial steels, such as are used on a large scale for constructional purposes, show ageing phenomena both after rapid cooling from high temperatures (quenching) and after mechanical deformation at room temperature. Among these phenomena there is a gradual but decidedly adverse changing of the mechanical properties of the steel, as for instance its hardness. When such changes take place after rapid cooling one speaks of "quench ageing", and in the other case of "strain ageing".

It will be made clear that the cause of this ageing in both cases lies in the presence of certain constituents in the steel. Of these, for our experiments and investigations, manganese, carbon, nitrogen and oxygen are of particular importance. These elements frequently occur in the steels in question in roughly the following weight percentages: 0.5 % Mn, 0.1% C, 0.01% N, 0.01% O. The statements to be found in the literature about the effect of each of these elements separately upon ageing contradict each other. This is due to the fact that the conclusions drawn are usually based upon experiments with the commercial steels themselves, that is to say with iron in which the said impurities and some others (particularly sulphur and phosphorus) are present simultaneously.

By employing the apparatus already described in this journal[1]), by means of which it is possible to prepare iron with an exactly known content of one or more impurities, we have been able to study separately the individual effects of carbon, nitrogen and oxygen in otherwise pure iron and in iron containing a little manganese.

In the course of the investigations dealt with in this article quench ageing was studied with the aid of hardness tests after various thermal treatments. New results, of importance for the theory of ageing and the technology of metals, were obtained by studying alloys containing nitrogen and others containing both manganese and nitrogen[2]).

The internal structure of iron and iron alloys

Before proceeding to discuss the experiments it is necessary to say something about the internal structure of the alloys, the degrees of solubility of the constituents mentioned, and the atomic processes in the metal which are to be held responsible for quench ageing. The subject matter of the following sections also forms the basis for the discussions in the second article.

Iron occurs in two modifications distinguished one from the other by the arrangement of their atoms. Below 910 °C and from 1400 °C up to its melting point (1540 °C) iron crystallizes into the body-centered cubic structure, a structure the smallest unit of which (the elementary cell from which an iron crystal of any size can be imagined to be built up by a continuous repetition in three mutually perpendicular directions) is illustrated in *fig. 1.* The cell contains only two atoms, since each of the eight atoms at the corners is common to eight cells. Below 910 °C this structure is denoted as α-iron and above 1400 °C as δ-iron. Iron in which the atoms are arranged in this manner is also sometimes called ferrite or ferritic[3]) iron, whilst iron alloys having this crystal structure are called ferritic alloys.

---

[1]) J. D. Fast, Philips techn. Rev. **11**, 241-244, 1949.

[2]) These results have been briefly reported in: J. D. Fast, Revue de Métallurgie **47**, 779-786, 1950 (No. 10).

Between 910 °C and 1400 °C iron crystallizes into a face-centered cubic structure, a structure whose elementary cell is depicted in *fig. 2*. This cell contains four atoms, since the atoms at the corners lie only for one-eighth part in the cube and those in the middle of the sides for only one half. In this case one speaks of γ-iron, or austenite, and of austenitic alloys.

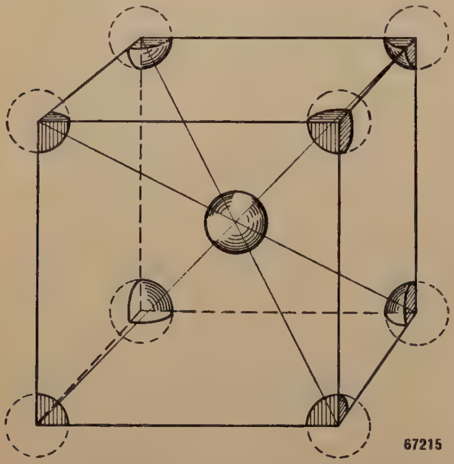


Fig. 1. Elementary cell of a body-centered cubic crystal, i.e. the structure of α-iron. In order to demonstrate that the elementary cell contains only two atoms, only those parts of the atoms at the corners are shown which lie within the cube.

In the case of a solution of manganese in solid iron (α- or γ-iron) one has to imagine iron atoms arbitrarily distributed over lattice sites as being replaced by manganese atoms. In a homogeneous austenitic alloy of iron and manganese with 25 atomic % Mn, each elementary cell (fig. 2) will therefore on an average contain one atom of Mn and three atoms of Fe. In such cases one speaks of substitutional solid solutions or "substitution mixed crystals".

When, however, elements such as carbon, nitrogen and oxygen, with small and comparatively readily deformable atoms, are present in solid iron as a solute, these atoms are located in the spaces between the iron atoms. Each elementary cell retains its normal number of iron atoms, but here and there one of them contains in addition an atom of the foreign element. In that case one speaks of interstitial solid solutions.

A closer investigation into the dimensions of the interstices available shows that both in α-iron and γ-iron there are two kinds of cavities to be distinguished. Supposing that the iron atoms in the crystal are touching spheres with radius R, then the radii of the largest spheres fitting in the cavities between the "iron spheres" have the following values:

in α-iron 0.29 R and 0.155 R,
in γ-iron 0.41 R and 0.23 R.

Thus the cavities in a body-centered crystal are much smaller than those in a face-centered crystal. However, the number of cavities in the α-structure is so much larger that the total space available in γ-iron is less than that in α-iron[4]). This is in agreement with the fact that the atoms in γ-iron form a close-packed structure, while the packing in α-iron is less dense.

The radii of the carbon and nitrogen atoms are much larger than those of the largest cavities in α-iron and γ-iron, amounting respectively to $r_C = 0.62\ R$ and $r_N = 0.56\ R$. The dissolution of carbon or nitrogen must, therefore, result in a considerable deformation of the elementary cell. This also holds when it is assumed that the carbon atoms are present in the lattice as monovalent or divalent positive ions and the nitrogen atoms as monovalent or divalent negative ions[5]).

When carbon or nitrogen is dissolved in γ-iron the deformation is equivalent with respect to the cube axes regardless of the kind of cavities occupied (there are indications that the carbon and nitrogen atoms occupy only one of the two kinds of cavities).

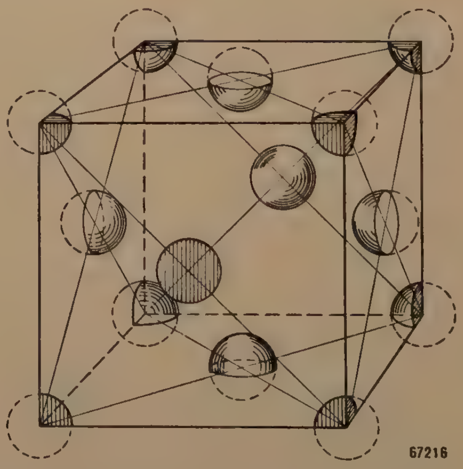


Fig. 2. Elementary cell of a face-centered cubic crystal. i.e. the structure of γ-iron. In this diagram and in fig. 1 the atoms have been drawn on too small a scale for the sake of clarity. To show that the elementary cell contains four atoms, only those parts of the atoms have been drawn which lie within the cell.

---

[3]) Not to be confused with compounds of the type $MeO.Fe_2O_3$ (in which Me represents a divalent metal), which have become known as ferromagnetic materials and are likewise called ferrites.

[4]) The α-structure contains six large (0.29 R) and three small (0.155 R) cavities per iron atom, the γ-structure only one large (0.41 R) and two small (0.23 R) cavities per iron atom.

[5]) Cf. W. Seith, Diffusion in Metallen, Springer, Berlin, 1939, pp 134 et seq.

In α-iron however the deformation is not symmetrical. This is illustrated in *fig. 3*. An atom situated at the point *1* (one of the smallest cavities available in the α-iron crystal) will bring about a tetragonal deformation of the elementary cell in the direction indicated in the diagram by y. By tetragonal distortion is to be understood such a distortion of the cube that one of the three mutually perpendicular edges is made longer or shorter than the others, as the case may be. An atom in the position *2* (one of the large cavities in the α-iron crystal) also causes a tetragonal but smaller deformation in the y direction. Atoms in the other available cavities will likewise cause tetragonal deformations but not necessarily always in the y direction; according to their position this deformation may take place in the x, y or z directions. It is therefore useful to distinguish between the so-called x, y and z positions, both for the large cavities and for the small ones; atoms in these positions will cause deformations in the x, y and z directions respectively.

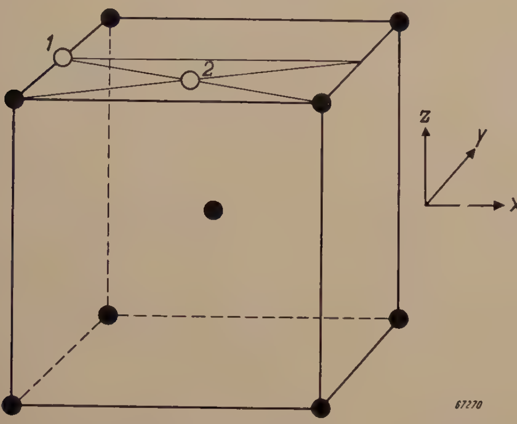


Fig. 3. The numbered positions are two so-called γ-cavities in the lattice of α-iron. The position *1* corresponds to the small cavities between the iron atoms, the position *2* corresponding to the larger ones. Foreign atoms at either of these positions cause a tetragonal deformation of the elementary cell in the y-direction.

In an equilibrium solution of carbon or of nitrogen in α-iron, not subject to any external forces, the solute atoms will be evenly distributed over the x, y and z positions. It must not be imagined that the C or N atoms are always in the same place; on the contrary they migrate from place to place through the crystal. There is, however, a random distribution at any moment, the lattice then being equally deformed in all directions and remaining cubic. A solid solution of carbon or nitrogen in α-iron having a tetragonal structure can, however, be obtained by quenching this C- or N-containing iron from the γ-phase. In the α-iron then obtained, the C or N atoms are found to be present mainly in interstices of only one of the three classes (x, y or z). Such tetragonal solutions are named martensite; they may be formed at the technical hardening of iron and steel. From now on we shall consider only quenching from the α-phase itself; the phenomena connected with technical hardening will not be discussed here.

Solubility of Mn, C, N and O in α-iron

The foregoing might create the impression that a metal like manganese, which with iron forms substitutional solid solutions, could do so when mixed in any proportions: in the picture given one could continue replacing Fe atoms by Mn atoms until pure manganese is obtained. Similarly one might gain the impression that the introduction of C or N atoms in, say, α-iron could be continued until all the available places are occupied. As the small cavities are filled this would lead to the formula FeC_3 or FeN_3 respectively, and as the large cavities are filled it would lead to FeC_6 and FeN_6 respectively [4]). This, however, is contrary to what is actually found in practice, which shows that Mn, C or N can be taken up in solid iron only in limited quantities. Predictions as to limits of solubility cannot be based upon the primitive geometrical considerations of the preceding section. They would have to be based upon atomic and thermodynamic calculations, but the theory of the solid state is not yet sufficiently developed to make such calculations possible [6]).

Thus one has to be guided almost entirely by experience, and this teaches that in the cases under consideration here, upon a certain quantity of the added element being exceeded, the temperature remaining the same, a second phase arises in the alloy with a composition differing from the first one. Crystals of this new composition are then formed at different places at the surface of the metal. The more of the foreign element is added, the larger these crystals become, until at last the whole metal is transformed into the new phase. The compositions of the phases in equilibrium with each other are, as a rule, greatly dependent upon the temperature. These relations can be read from diagrams called equilibrium or phase diagrams.

As an example *fig. 4* shows the part of the phase diagram for iron-manganese which is of interest to us here. The lines in this diagram indicate the limits within which different phases are stable. Along the temperature axis the composition is that of pure iron, with as characteristic point the transition, already mentioned, at 910 °C. As is seen, when manganese is added the transition α-γ is shifted to lower temperatures. It is not our intention to go into all details of the diagram, but it should be pointed out that — as is always the case in such diagrams — the domains within which only one phase occurs are separated by a region in which the two separate phases are co-existent. As the diagram shows, the solubility of manganese in α-iron at room temperature is at most about 3%; when more manganese is added the γ-phase and the α-phase exist at the same time.

---

[6]) The same applies, in fact, for answering the obvious question as to why below 910 °C and above 1400 °C iron crystallizes in the body-centered cubic structure, whereas between 910 °C and 1400 °C the structure is face-centered cubic.

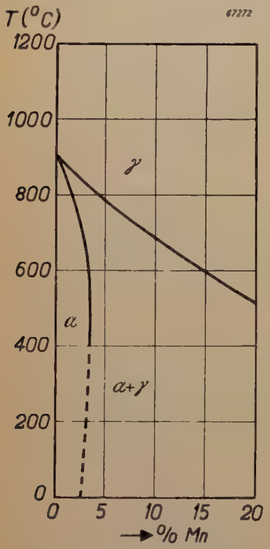


Fig. 4

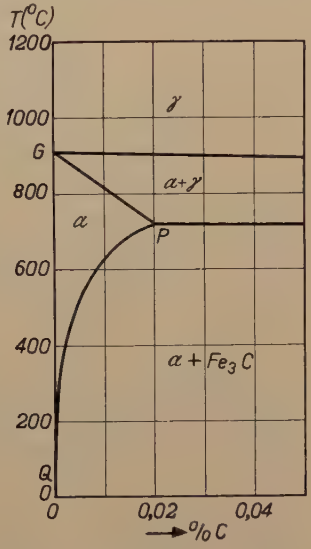


Fig. 5

Fig. 4. Part of the phase diagram for iron-manganese. The zone above 1200 °C (where, i.a., the γ-δ transition takes place and the liquidus line is situated) has not been drawn. (Taken from A. R. Troiano and F. T. McGuire, Trans. Am. Soc. Met. **31**, 340-364, 1943.)

Fig. 5. Part of the phase diagram for iron-carbon. The zone above 1200 °C has not been drawn. The line PQ represents the equilibrium between α-iron with carbon in solid solution and iron carbide. This diagram clearly shows that by rapid cooling from the α-phase a supersaturated solution of carbon in α-iron can be obtained. From this solution there will be a gradual precipitation of iron carbide.

The solubilities of carbon, nitrogen and oxygen in α-iron are much less than that of manganese. For carbon this is demonstrated in *fig. 5*, representing part of the equilibrium diagram Fe-C. Just as in the case of manangese, the solubility in the γ-phase is much greater than that in the α-phase.

If the α-iron contains more carbon than corresponds to the solubility line PQ then, in the state of equilibrium, in addition to the saturated solution there is a quantity of iron carbide (Fe_3C; cementite) present [7].

The curve GP, covering the temperature range between 910 °C and 720 °C, is the solubility curve of carbon in α-iron in equilibrium with carbon in γ-iron. The maximum quantity of carbon that can be dissolved in α-iron is reached at the so-called eutectoid temperature of 720 °C, where the two solubility curves coincide.

In the region of small nitrogen contents the iron-nitrogen diagram has an appearance similar to the iron-carbon diagram, but the eutectoid temperature lies lower, viz. at 580 °C. Below this temperature, upon the solubility limit being exceeded, the nitride Fe_4N occurs as a second phase [8]).

Only recently the position of the line PQ became known with any accuracy [9]), the solubilities of C and N, expressed in weight %, being represented by the formulae:

$$\log (\%C) = -\frac{2200}{T} + 0.50 \quad \text{and}$$

$$\log (\%N) = -\frac{1550}{T} + 0.70,$$

where T is the absolute temperature. These formulae are based upon measurements between 200 °C and the eutectoid temperatures. They show that the solubility of nitrogen is much greater than that of carbon.

For the solubility of oxygen in α-iron the literature on the subject gives greatly conflicting values. From the fact, however, that in iron containing oxygen the same transitional points are found as in pure iron, and from our own experiments on the ageing of oxygen-containing iron it can be concluded that the solubility of oxygen is almost zero and in any case much less than that of carbon and nitrogen. Even at high temperatures, nearly all the oxygen in oxygen-containing iron is thus present in the form of oxide in solid iron.

Nucleation, diffusion and precipitation

In the foregoing it has been seen that below the eutectoid temperature the solubility of carbon and nitrogen in α-iron decreases with the temperature (see the line PQ in fig. 5 and the formulae given above). When, therefore, carbon- or nitrogen-containing α-iron is cooled from a high temperature, according to the phase diagram a precipitation of carbide, or nitride, respectively, may be expected. For such a precipitation to take place, first nuclei of the second phase have to be formed in the homogeneous solution. These nuclei will then grow, because the carbon or nitrogen atoms diffuse through the solution towards them.

---

[7]) Strictly speaking, in the state of equilibrium graphite is present as second phase. Due, however, to nucleation difficulties this completely stable state of equilibrium is not reached, but rather the metastable state mentioned.

[8]) Here, too, strictly speaking, the state of equilibrium is metastable, since at the temperatures in question and under a pressure of 1 atm Fe_4N should spontaneously dissociate into iron and nitrogen, which, however, does not take place at a perceptible rate.

[9]) L. J. Dijkstra, Trans, A.I.M.E. **185**, 252-260, 1949.

Now the processes of nucleation and of diffusion both take some time. If the cooling, say from the eutectoid temperature to room temperature, takes place quickly, then supersaturated solutions will be obtained. As a rule, it is difficult for the nuclei to form when cooling is stopped at comparatively high temperatures, whilst diffusion takes place very slowly at comparatively low temperatures. Thus the rate at which the total process of precipitation takes place may be determined by the rate of nucleation at relatively high temperatures and by the rate of diffusion at relatively low temperatures. In the case now under consideration (C or N in α-iron), at room temperature and at temperatures not much higher it is in all probability diffusion that determines the rate of the total process of precipitation.

From the foregoing considerations, quench ageing — the change in the hardness of the metal which is apt to take place in commercial steel rapidly cooled from the α-phase — can be correlated with a phenomenon of precipitation, namely the precipitation of carbide or nitride, from a supersaturated solution of carbon or nitrogen in α-iron[10]). It should be borne in mind that the plastic deformation of a metal is brought about owing to parts of the metal crystals slipping over each other along certain crystallographic planes. The resistance of the metal to slip increases by the presence of the interstitial solute atoms. There is, however, a much greater increase in slip resistance and thus a much greater increase in hardness if a precipitate is present in the form of a large number of minute carbide or nitride particles in the metal crystals.

If the precipitation is made to take place at a constant temperature (say 100 °C), first the hardness is seen to increase gradually and then to decrease again very gradually. Such a change occurs also in the case of the frequently investigated substitutional solid solutions of copper in aluminium. In an attempt to explain this, in 1919 Merica, Waltenberg and Scott assumed that first a very fine precipitate of a Cu-Al compound (Al_2Cu) is formed, which gradually increases in size. The maximum hardness would be reached at a "critical degree of dispersion" of the Al_2Cu lying below microscopic resolution. Later it appeared that the process of precipitation in copper-containing aluminium is much more complicated: there is not a decidedly heterogeneous (two-phase) system right from the beginning of the precipitation, as in the old picture, but rather a system in which the degree of heterogeneity increases gradually, with intermediate stages where there is neither a distinct homogeneous nor a distinct heterogeneous system[11]).

There are no reasons to assume, however, that similar intermediate stages occur in the process of precipitation taking place in the simple interstitial alloys Fe-C and Fe-N (at least if these alloys have not previously been plastically deformed). With these alloys the precipitation seems to follow a simpler course, such that a second phase is at once formed in the shape of minute carbide or nitride crystals, but with this complication (see[9])) that within a certain temperature range in the Fe-N system a nitride phase is first formed (probably $Fe_{16}N_2$[12])) which is less stable than Fe_4N. Nuclei of $Fe_{16}N_2$ are formed more readily and thus in larger numbers than those of Fe_4N. Most of the nitrogen, therefore, begins to precipitate in the form of $Fe_{16}N_2$, which, however, gradually dissolves as the more stable Fe_4N nuclei increase in number.

In the Fe-C system it seems that in precipitation Fe_3C alone occurs as second phase. In this case, too, however, an equilibrium state is not yet reached when the supersaturation of the solid solution has ceased by the formation of Fe_3C; some Fe_3C crystals will gradually grow at the expense of others, which again dissolve because the total interfacial energy is thereby decreased. Whereas the precipitation in its first stages causes an increase in hardness and a reduced malleability, as the precipitate becomes coarser — the process being called "coagulation" (occurring also in the case of the nitride) — hardness again decreases and malleability increases. It is difficult, however, in this connection to ascribe any sharply defined significance to the "critical degree of dispersion" of Merica c.s., since at any moment there will always be a range of particle dimensions owing to the simultaneous growth of some carbide particles and the dissolution of others (which in fact already takes place before the supersaturation is entirely overcome).

Experiments on ageing

In order to study the influence of each alloy element separately on quench ageing, alloys of exactly known compositions were prepared in the apparatus already referred to[1]). From these alloys flat test bars about 4 mm thick were forged, these then being heated to 940 °C and slowly cooled before the tests were begun. This took place in a protective atmosphere to prevent change in composition of the bars.

---

[10]) This phenomenon was first pointed out by Köster in 1929; see W. Köster, Arch. Eisenhüttenwesen **2**, 503-522, 1928-'29; **3**, 553-558, and 637-658, 1929-'30.

[11]) Cf. e.g. A. Guinier, Physica **15**, 148-160, 1949 and G. D. Preston, Phil. Mag. **26**, 855-871, 1938.

[12]) K. H. Jack, Acta Cryst. **3**, 392-394, 1950 (No. 5).

The bars were then heated for some time in vacuo at the quenching temperature, this temperature being chosen for each alloy to be slightly below its eutectoid temperature [13]). After this heating the bars were rapidly cooled by immersion in a large quantity of cold water. Immediately after quenching the Vickers hardness was measured at a temperature of 0 °C. This hardness value is measured by pressing with a certain load a quadrilateral pyramid of diamond into a smooth-ground surface of the metal. The impression left in the surface of the metal is square-shaped, and the diagonals of the square are measured. By the term hardness H is understood the force in kilogrammes divided by the pyramidal area (the sum of the areas of the four sides) of the indentation in mm^2.

After this measurement had been taken at 0 °C the hardness was measured again a number of times at 20 °C, namely after two hours' heating at 50 °C (followed by slow cooling), then after two hours' heating at 100 °C, then again after two hours' heating at 200 °C, and so on. In this way a picture of the phenomenon of precipitation can be obtained much more quickly than by measuring the hardness as a function of time at each of these temperatures, though in principle this is the most elegant method.

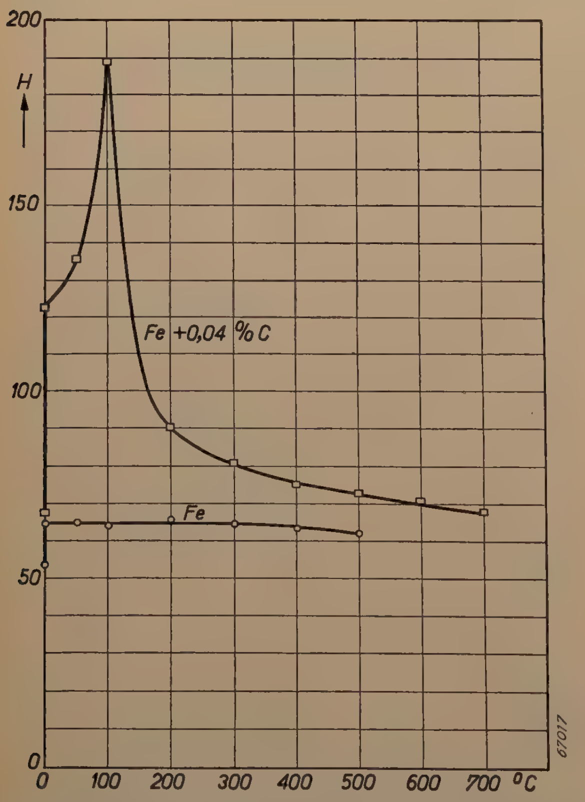


Fig. 6. Changes taking place in the Vickers hardness value H of iron with 0.04% carbon (squares) and of pure iron (circles) after the heat treatments described in the text.

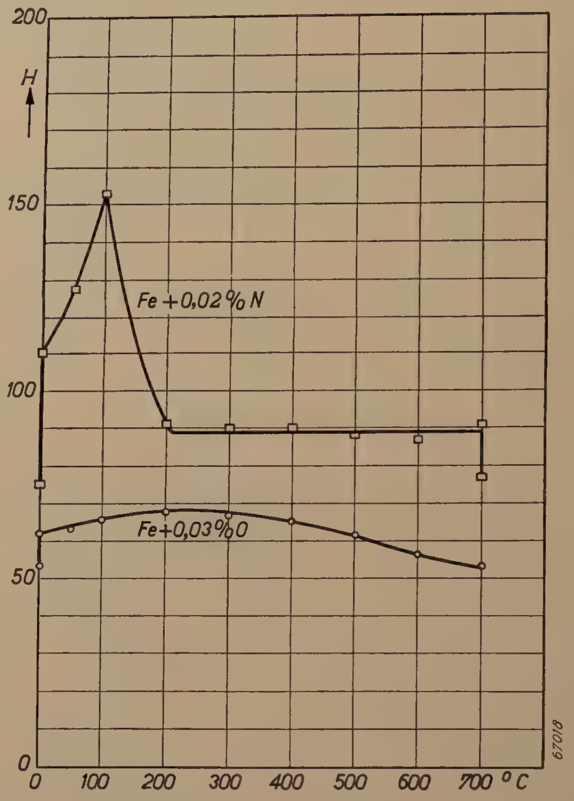


Fig. 7. As in fig. 6 but for iron with 0.02% nitrogen (squares) and for iron with 0.03% oxygen (circles). The lowest hardness value on the ordinate for 700 °C of nitrogen-containing iron was reached after re-heating to 700 °C and then very slowly cooling (during some hours) from that temperature. This further decrease in hardness occurs only with nitrogen-containing iron.

As already stated, the general trend, at a constant temperature, in the change of the hardness of a material that shows quench ageing is characterized by first a rather rapid increase followed by a slow decrease, until approximately the same hardness is reached which the metal possessed prior to the quenching. To be able to observe the slow change within a limited space of time, the bars have to be heated to a relatively high temperature, because the higher the temperature the more quickly the process takes place, as will be evident from what has been said above. The method of working described is based on this fact alone and has no further physical significance.

The materials used for these experiments were pure iron, iron with 0.03% oxygen, iron with 0.02% nitrogen, iron with 0.04% carbon and, further, the last two alloys with an extra admixture of 0.5% Mn. The results are shown graphically in *figures 6, 7* and *8*.

---

[13]) The iron-oxygen system does not show any eutectoid point, or at least not one lying noticeably below the transition point of pure iron. In this case quenching was done from 750 °C.

In addition to the hardness values determined after the various heat treatments described above, the graphs also show the hardness values of the alloys prior to quenching, which are lower.

As was expected, pure iron (fig. 6) shows no quench ageing. The quenching does, it is true, set up internal stresses in the metal which cause a slight increase in hardness, but the heat treatments at 50 °C, 100 °C and higher cause practically no further increase. The picture is quite different with the iron containing only carbon and that containing only nitrogen (figs 6 and 7). Owing to the dissolution of the carbon and nitrogen, the hardness after heating to the eutectoid temperature and quenching is greater than that prior to quenching. Under the heat treatments following upon quenching there is, however, a further considerable increase in hardness, the maximum being reached after heating to 100 °C. This temperature has no physical significance, on account of the arbitrarily chosen treatment to which the material was subjected. Oxygen-containing iron (fig. 7) shows only very little quench ageing.

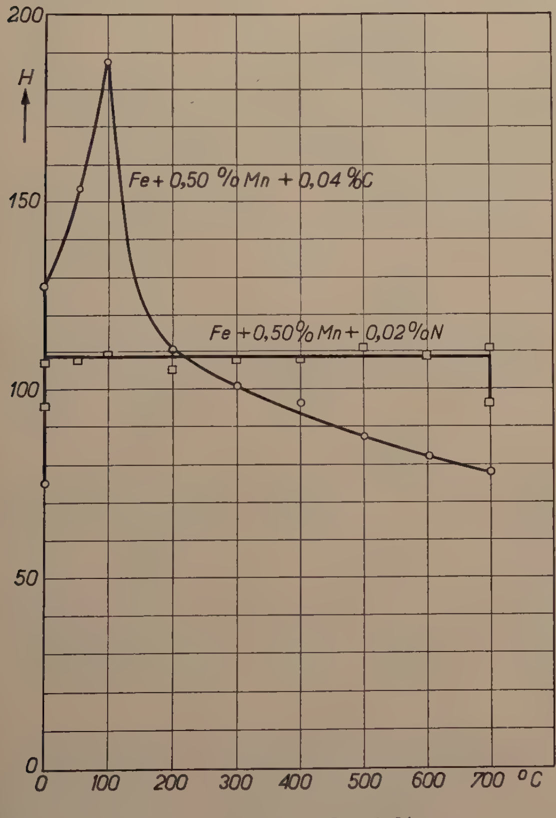


Fig. 8. As in fig. 6 but for iron with 0.50% manganese and for iron with 0.04% carbon and 0.02% nitrogen. The quench ageing of nitrogen-containing iron is suppressed by the manganese. Here again there is a further decrease in hardness of nitrogen-containing iron after slow cooling from 700 °C.

The experiments show that the presence both of carbon and of nitrogen may lead to phenomena of precipitation in iron. The fact that oxygen has so little effect has to be ascribed to the extremely small degree of solubility of this element in α-iron, even at such a high temperature as 750 °C.

A very important result of the experiments is that manganese appears to have but little influence upon the precipitation behaviour of carbon-containing iron (cf. figs 6 and 8), but that this element tends to suppress the precipitation in nitrogen-containing iron (cf. figs 7 and 8).

The fact that in the presence of 0.5% manganese no perceptible quench ageing occurs in iron containing nitrogen may be accounted for in the following way. It is presumed that the nitrogen atoms are preferably present in such interstitial sites that at least one of their immediate neighbours is a manganese atom, for the reason that in such places a nitrogen atom will possess less energy than when it has exclusively iron atoms for its immediate neighbours. This is quite feasible considering that manganese and nitrogen. show a greater chemical affinity than iron and nitrogen. Consequently for the nitrogen atoms to break away from the manganese atoms and then to be able to precipitate, the temperature has to be so high that coagulation also proceeds at a fast rate. Thus there would not be any perceptible temporary increase of the hardness.

It is possible to put the foregoing supposition to the test by means of damping experiments. These will be dealt with in the following article.

Summary. After quenching (rapid cooling) of iron or steel a slow change often takes place in the mechanical properties of the material. This quench ageing is to be ascribed to the presence of certain constituents in the metal. The hardness of α-iron containing small quantities of carbon, nitrogen or oxygen separately has been determined after various heat treatments. Also the influence of manganese on the ageing of α-iron with carbon or nitrogen has been studied. This investigation has been confined to a study of the changes in hardness taking place after quenching from the α-phase, thus from a temperature lying below the eutectoid temperatures of the materials; the technically still more important case of quenching from the γ-phase has been left out of consideration. It appears that oxygen causes virtually no quench ageing; this element is, in fact, almost insoluble in α-iron. Carbon and nitrogen are to some extent soluble in α-iron, and both cause considerable quench ageing. This is to be ascribed to precipitation of carbon and nitrogen in the form of carbide and nitride, from their supersaturated solution in the iron formed by rapid cooling. Manganese appears to be capable of suppressing the quench ageing of iron containing nitrogen but not that of iron containing carbon. Considering the comparatively great affinity between manganese and nitrogen, an explanation is sought by assuming that in the crystal lattice the nitrogen atoms will preferably occur in the vicinity of manganese atoms, where they are bound much more strongly than when surrounded exclusively by iron atoms, so that the formation of iron nitride, thus the precipitation, takes place at a perceptible rate only at high temperatures.

INTERNAL FRICTION IN IRON AND STEEL

by J. D. FAST and L. J. DIJKSTRA.

539.312 : 534.372 :
620.193.918 : 669.14-156

The internal friction of torsional oscillations of small amplitude of iron and steel wires is an after-effect phenomenon which under certain conditions arises from the diffusion of foreign atoms in the iron lattice. An application of this phenomenon is to give an insight into the precipitation of carbon or of nitrogen in α-iron, which precipitation — as shown in the preceding article — is responsible for quench ageing. The damping experiments which have been carried out throw more light upon the part played in this connection by manganese in nitrogen-containing iron.

Introduction

In the preceding article [1]) some phenomena were discussed which are liable to take place in iron and manganese-containing iron when carbon or nitrogen are also present. Following upon those discussions, an account will be given of some experiments which have been carried out with a view to obtaining a better insight into the atomic background of these phenomena. The experiments are based upon the fact, now known for about ten years, that the presence of carbon or nitrogen in iron may lead to a strong "internal friction", i.e. a rapid decay of mechanical vibrations set up in the metal.

Before discussing the experiments and the interpretation of their results, it is necessary to consider briefly the more widely known phenomenon of "external friction" and then to explain how it is possible for the presence of carbon or nitrogen *in* the metal to have the same effect upon mechanical vibrations as that of friction with a medium outside the metal. It will then appear, inter alia, that carbon and nitrogen can cause friction only if they are present in the iron in the solute state, and not if they are present in the form of carbide and nitride.

Torsional oscillations and external friction

In our damping experiments use was made of wires of the alloys to be investigated. Such a wire formed the elastic element of a torsional pendulum; its upper end was tightly clamped, while a body with a known moment of inertia was suspended from the other end. When the free end is twisted over a certain angle and then released the wire starts to make torsional oscillations. The potential energy of the twisted wire is converted into kinetic energy of the inert mass of the body and vice versa. The deformation of the wire is determined by an elastic couple K directed opposite to the deflection, which couple in the case of a small deflection is proportional to the angle of deflection φ:

$$K = -a\varphi,$$

where a is called the rigidity with respect to torsion or, briefly, the torsional rigidity of the wire. According to Newton's equations of motion this couple should always equal the product of the moment of inertia I of the pendulum and the angular acceleration $\mathrm{d}^2\varphi/\mathrm{d}t^2$. Thus we have the relation:

$$I\frac{\mathrm{d}^2\varphi}{\mathrm{d}t^2} = -a\varphi. \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

In our case the solution of this differential equation is:

$$\varphi = \varphi_0 \cos \omega t, \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

and hence:

$$K = -a\varphi_0 \cos \omega t = a\varphi_0 \cos (\omega t + \pi),$$

provided the zero point of the time is fixed at a moment when the angle φ has the maximum value φ_0. The angular frequency, ω, is given by:

$$\omega = 2\pi\nu = \frac{2\pi}{T} = \sqrt{\frac{a}{I}}, \quad . \quad . \quad . \quad . \quad . \quad (3)$$

where ν is the frequency and T the periodic time of the oscillations.

According to eq. (2) the torsional pendulum would continue to oscillate for an infinitely long time. In reality such oscillations always decrease in amplitude, i.e. free vibrations are always damped. The best known cause of this is external damping, due to friction between the oscillating body and the surrounding medium (e.g. air). The frictional couple is always opposed to the angular velocity and for not too high velocities is proportional to it; it can therefore be expressed by the formula:

$$W = -c\frac{\mathrm{d}\varphi}{\mathrm{d}t}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

---

[1]) J. D. Fast, Ageing phenomena in iron and steel after rapid cooling, Philips techn. Rev. 13, 165-171, 1951 (No. 6), hereinafter referred to as article I.

Provided the friction is only small, the decrease in amplitude due to such a frictional couple follows an exponential law; in this case the relation between the angle of deflection and the time is represented to a good approximation by:

$$\varphi = \varphi_0\, \mathrm{e}^{-\frac{\gamma}{2}t} \cos \omega t, \quad . \; . \; . \; . \; . \; . \; . \; . \quad (5)$$

in which γ is given by:

$$\gamma = \frac{c}{I}. \quad . \; . \; . \; . \; . \; . \; . \; . \; . \; . \; . \quad (6)$$

The zero point of the time lies at the instant when, after an angular twist $\varphi = \varphi_0$ is given to the system, it is left to itself.

The extent to which the amplitudes of vibration decrease with time can be expressed in various ways. A quantity frequently employed is the logarithmic decrement δ, this being the natural logarithm of the ratio of two amplitudes φ_1 and φ_2 lying one periodic time T apart. From eq. (5) we find for δ in our case:

$$\delta = \log_e \frac{\varphi_1}{\varphi_2} = \log_e \frac{\mathrm{e}^{-\frac{\gamma}{2}t} \cos \omega t}{\mathrm{e}^{-\frac{\gamma}{2}(t+T)} \cos \omega\,(t+T)} = \gamma \frac{T}{2} = \pi \frac{\gamma}{\omega}. \quad (7)$$

Another method of expressing that energy is dissipated, thus that damping occurs, is to state that the angle of deflection φ and the total counteracting couple V—now the sum of the elastic couple K and the frictional couple W— are no longer exactly in anti-phase, or, in other words, that they no longer show a phase difference π.

The deviation from π equals an angle β, the loss angle, which, in the case of not too great a friction, is given by:

$$\beta = \frac{\gamma}{\omega} = \frac{\delta}{\pi} \quad . \; . \; . \; . \; . \; . \; . \; . \quad (8)$$

(all angles are expressed in radians).

The above results follow from the differential equation for the damped free vibration, which reads:

$$I\,\frac{\mathrm{d}^2\varphi}{\mathrm{d}t^2} = -a\varphi - c\,\frac{\mathrm{d}\varphi}{\mathrm{d}t}, \quad . \; . \; . \; . \; . \; . \quad (9)$$

as becomes obvious from the text (formulae (1) and (4)).

In the case that $\gamma = c/I \ll \omega$ (the angular frequency of the undamped vibration), to a very good approximation the solution of this equation is the expression given in (5).

The couple V opposing the motion is given, according to (9), by:

$$V = -a\varphi - c\,\frac{\mathrm{d}\varphi}{\mathrm{d}t}.$$

After some calculation, employing the relations (3) and (6) and the condition $\gamma \ll \omega$, substitution of (5) in the formula for V gives:

$$V = -a\varphi_0\, \mathrm{e}^{-\frac{\gamma}{2}t} \left(\cos \omega t - \frac{\gamma}{\omega} \sin \omega t\right),$$

which expression, again because $\gamma \ll \omega$, can to a good approximation be written as:

$$V = a\varphi_0\, \mathrm{e}^{-\frac{\gamma}{2}t} \cos\,(\omega t + \pi + \beta).$$

Here the loss angle β is given by:

$$\sin \beta \approx \beta = \frac{\gamma}{\omega},$$

which, according to (7), is equal to δ/π.

Thus the couple V differs an angle $\pi + \beta$ in phase from the angle of deflection φ, thereby proving the statement set forth in the foregoing.

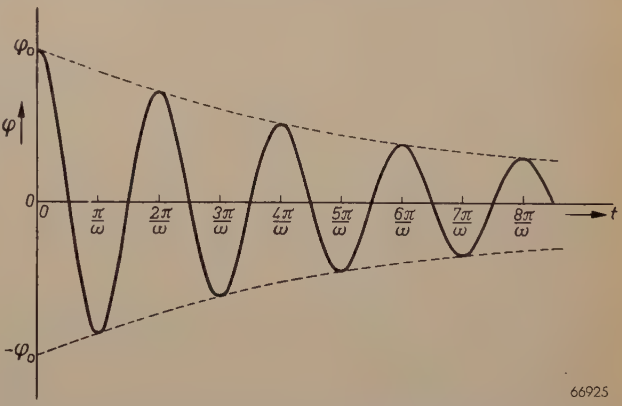


Fig. 1. Torsional oscillations with angular frequency ω decay exponentially when they are externally damped. The enveloping curves $\varphi = \varphi_0\, \mathrm{e}^{-\frac{\gamma}{2}t}$ and $\varphi = -\varphi_0\, \mathrm{e}^{-\frac{\gamma}{2}t}$ are given in broken lines. The points where these enveloping curves coincide with the fullydrawn curve correspond almost to the maximum and minimum values of φ. It will be shown that a similar relation exists between φ and t for the internal friction. This diagram applies for a value of $\gamma/\omega = 0.1$. In practice the damping is mostly much less than this.

In *fig. 1* it is illustrated how the amplitude of the damped oscillations decreases with time. It demonstrates the significance of the condition $\gamma/\omega \ll 1$ (viz. that the relative decrease of the amplitude during one periodic time is small).

After-effect, relaxation and phase shift

To understand how a vibration can be damped also as a result of the presence of dissolved carbon or nitrogen (thus, how vibrational energy can be converted into heat by internal causes), we have to start from the fact, discussed in article I, that the deformation caused by interstitially dissolved atoms of these elements in α-iron is not uniform with respect to the axes of the cube. A distinction was made between x, y and z positions. Carbon or nitrogen atoms in x positions cause a tetragonal deformation in the x direction, those in y positions a similar deformation in the y direction, and so on.

Let us now imagine a single crystal of α-iron in the shape of a cube in which a quantity of carbon is present in homogeneous solution *(fig. 2)*. So long as the cube is not subjectief to external forces, the carbon atoms are uniformly distributed among the x, y and z positions. If, however, the cube is elongated elastically in the z direction (which is assumed to coincide with the z directions of the elementary cubes in the crystal) then in the z positions more space becomes available for the carbon atoms, whereas, owing to the Poisson contraction, in the x and y positions there will be less space. Owing to this deformation the positions are no longer energetically equivalent: the carbon atoms will show a preference for the z positions. Whereas prior to the elastic deformation the time that the carbon atoms stay in any one interstice was on an average the same for all interstices, after the deformation this mean time of stay will be greater for the z positions than for the others. In other words, under the influence of the elastic deformation the diffusion equilibrium is shifted in such a way that more z positions and less x and y positions are occupied.

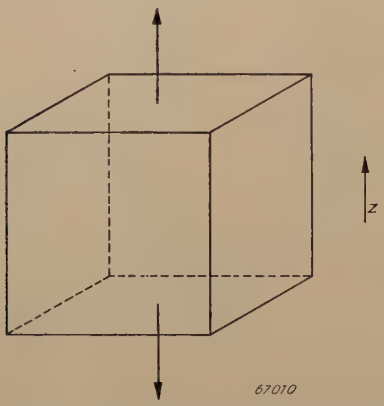


Fig. 2. A single crystal of α-iron in the form of a cube stretched in the z direction; as a result there will be some contraction in the x and y directions.

It is clear that this increased occupation of the z positions will lead to relaxation of the elastic force in the case of a constant deformation. When a constant load is applied the body will at once respond with a certain deformation, but under the influence of the displacement of the carbon atoms this deformation will gradually increase. Upon the load being removed the reverse takes place: part of the deformation disappears immediately, but the remaining deformation requires some time to do so, because it takes some time for the diffusion equilibrium to be restored. In the case under review one speaks of elastic after-effect.

The fact that carbon and nitrogen may, in the manner described, lead to after-effect phenomena in α-iron was first recognized by Snoek[2]). (Other possible causes of after-effect, such as the thermoelastic effect, will not be considered here.)

The foregoing considerations were related to a single crystal, whereas our experiments were carried out with polycrystalline wires. In a single crystal elongation in a direction parallel to the axes of the cube results in relaxation, but elongation in the direction of a solid diagonal of the cubic elementary cell has, of course, no such result. Distortion of a polycrystalline material in which the crystals have all possible orientations will therefore result in a relaxation lying somewhere in between the maximum value and zero.

It will now be investigated in what manner this relaxation affects torsional oscillations. As is known from the theory of elasticity, the torsion of an iron wire results in a compression of the volume elements of the iron in one direction (at an angle of 45° to the longitudinal axis) and an equally large elongation in a direction at right angles to the first. Thus also due to the torsion the diffusion equilibrium is shifted. Upon the direction of torsion being reversed the compressions change into elongations and the elongations into compressions. If the wire is twisted suddenly over a certain angle and that angle is thereafter kept contant, then the resultant elastic couple will gradually decrease owing to displacement of the carbon and nitrogen atoms. This is illustrated by *fig. 3*, where the rectangular line drawn in *a*

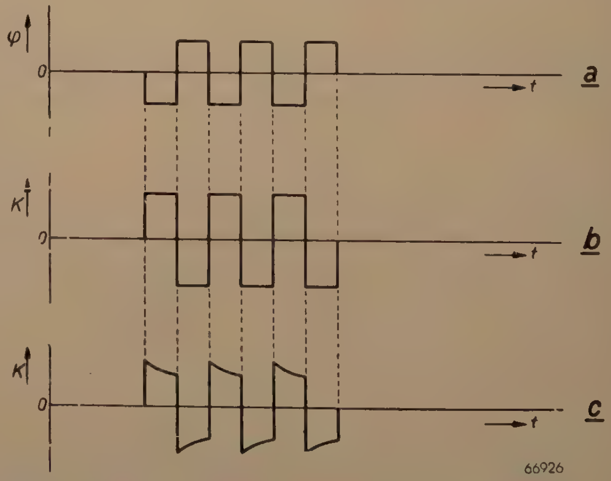


Fig. 3. When the angular deflection φ of a torsional pendulum varies discontinuously in the manner indicated (a) a medium without after-effect will react to this with an elastic counteracting couple K in the form of the curve (b). In a medium with after-effect the couple will vary as a function of time according to curve (c). During each period of a constant angular deflection the absolute value of the couple decreases. This decrease takes place practically exponentially.

---

[2]) J. L. Snoek, Physica **9**, 711-733, 1941; **8**, 862-864, 1942.

represents an angular deflection kept constant for a very definite time and then changed in sign in an immeasurably short time. A material without after-effect responds to this with a counteracting couple of the form represented in *b*, whilst a medium having after-effect reacts according to *c*.

For a relaxation based on a simple diffusion phenomenon as that described here it may be assumed that the couple K required for a constant angular deflection decreases with time such that $\mathrm{d}K/\mathrm{d}t$ is always proportional to the instantaneous value of the variable part of K. Thus it may be assumed that

$$-\tau \frac{\mathrm{d}K}{\mathrm{d}t} = K - K_e \,, \quad . \; . \; . \; . \; . \quad (10)$$

where τ is a proportionality constant with a time dimension, which is called the relaxation time of the elastic after-effect. If $K = K_b$ at the instant $t = 0$ then integration of (10) yields the relation:

$$K = K_e + (K_b - K_e)\, \mathrm{e}^{-\frac{t}{\tau}}. \quad . \; . \; . \quad (11)$$

K therefore decreases exponentially with time to a final value K_e which would be reached after an "infinitely long time". The relaxation time τ denotes how quickly this decrease takes place: after τ sec the difference between K and its final value has been reduced to the fraction $1/\mathrm{e} = 37\%$, after 2τ sec it has been reduced to $1/\mathrm{e}^2 = 13\%$, and so on.

If the angular deflection is made to reverse in sign after intervals of time which are very short compared with τ then the relaxation will have only very little effect; during the short periods of constant angular deflection the couple retains practically the value K_b (*fig. 4a*). In this case, therefore, one has to consider exclusively the relatively high initial value of the torsional rigidity, the "unrelaxed torsional rigidity". When, on the other hand, the angular deflection is made to change in sign after intervals of time which are long compared with τ, then during the long periods of constant angular deflection the couple has an average value very close to K_e (figure 4*b*). In that case one has to do with a smaller value of the torsional rigidity, the "relaxed torsional rigidity", corresponding to K_e.

Similar considerations hold when, instead of discontinuously as represented in figs 3 and 4, the value of φ changes gradually, for instance periodically according to a sine curve with angular frequency ω, as is the case with torsional oscillations. If the periodic time of the oscillations is short compared with the relaxation time, in other words if $\omega\tau \gg 1$, then the deformation is continuously changing sign so quickly that there is no perceptible displacement of the diffusion equilibrium and thus no perceptible relaxation; one has to consider the unrelaxed torsional rigidity. If, on the other hand, $\omega\tau \ll 1$ the diffusion equilibrium has every opportunity to adapt itself to the changing conditions and the wire behaves as a less rigid medium than in the first case: the relaxed torsional rigidity is then measured. In both cases the counteracting couple is continuously in anti-phase with the deformation.

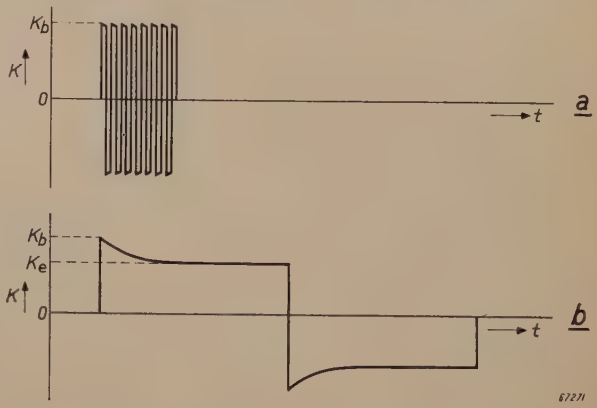


Fig. 4. *a*) When the angular deflection periodically alternates in sign at a frequency largely exceeding $1/\tau$, then during the short periods of constant angular deflection the counteracting couple K has a virtually constant value determined by the unrelaxed torsional rigidity. *b*) If the frequency is low with respect to $1/\tau$ then the mean absolute value of K during the moments of constant angular deflection is much smaller; one then has to do with the relaxed torsional rigidity.

In the frequency range for which $\omega\tau \approx 1$, that is to say in the range where the periodic time of the oscillations is of the same order as the time (a few times τ) in which the diffusion equilibrium can be almost entirely established, the influence of the interstitial solute atoms is quite different. To investigate this influence let it be assumed that the cube in fig. 2 is periodically compressed and elongated in the z direction. Within the range under consideration ($\omega\tau \approx 1$) the frequency of this compressing and elongating is not high enough to prevent a periodical variation in the occupation of the z cavities, but on the other hand it is not low enough to prevent a phase difference arising between the degree of occupation of the interstices and the deformation. The deviation from the average occupation becomes zero later and thus the counteracting force becomes zero earlier than the deformation.

Applied to our torsional oscillations this means that, apart from the difference in sign, the curve representing the variation of the couple is displaced over a certain distance β/ω, depending upon $\omega\tau$, with respect to the sine curve denoting the variation of the angular deflection. This is illustrated in *fig.* 5, where φ indicates the periodical angular deflection and K the corresponding counter-acting couple for a medium with relaxation.

If the angle of deflection is given by $\varphi = \varphi_0 \cos \omega t$, then for K the equation reads:

$$K = K_0 \cos (\omega\tau + \pi + \beta).$$

It has already been seen that the loss angle β occurring in this expression is small both for $\omega\tau \gg 1$ and for $\omega\tau \ll 1$. The variation of β in the intermediate range is represented in *fig. 6:* a maximum is reached where $\omega\tau = 1$, the level of this maximum being determined by the magnitude of the after-effect. The same diagram shows the variation of K_0/φ_0, which factor gradually changes with rising $\omega\tau$ from the relaxed to the unrelaxed torsional rigidity.

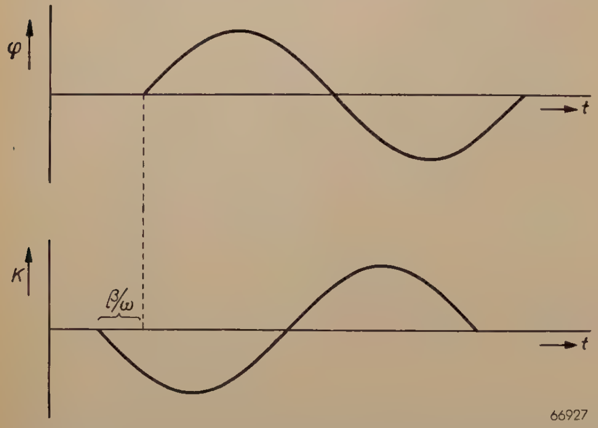


Fig. 5. When the relaxation time τ of the medium and the angular frequency ω of the pendulum satisfy the condition $\omega\tau \approx 1$ there is a phase difference $\neq \pi$ between the opposing couple K and the angle of deflection φ. The couple is a time $(\pi + \beta)/\omega$ in advance of this angle.

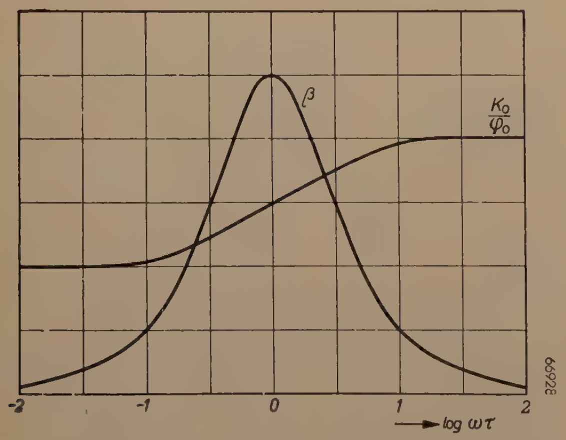


Fig. 6. β and K_0/φ_0 (the relation between the greatest values of the opposing couple and the angle of deflection) as functions of log $\omega\tau$.

It may sometimes happen that there is more than one relaxation time, for instance when different processes of diffusion are playing a part simultaneously. In that case the curve for β is a sum of two or more curves, each with its maximum at the value of ω corresponding to the respective relaxation time. If there are not too great differences between the relaxation times this superposition means a widening of the maximum in the curve for β.

Internal friction. Influence of temperature

From the discussion of external friction it has been seen that the occurrence of a phase shift β is accompanied by a dissipation of energy, thus, in the case of free oscillations, by a decay of the oscillations. This is the general rule, no matter what the cause of the phase shift may be. Therefore the loss angle β caused by the jumping of the carbon or nitrogen atoms from one interstitial position to another also leads to damping, which is denoted as "internal friction" as distinguished from the external friction dealt with in the beginning of this article. For $\omega\tau \gg 1$ and for $\omega\tau \ll 1$ the value of β was very small, which means that the jumping of the carbon or nitrogen atoms in these frequency ranges does not cause any damping. In the range where $\omega\tau \approx 1$, however, the internal friction in the case of our torsional pendulum is far from negligible, being even much greater than the external friction. In what follows, the latter will therefore be left out of consideration.

In absolute value the loss angle β determining the internal friction is also small for $\omega\tau = 1$, being at most a few degrees (the phase shift which determines the external friction is even smaller). It is therefore permissible to apply the arguments given in the beginning of this article; we then find that the amplitude of the free oscillations decreases exponentially. The order of the decrease is determined by the logarithmic decrement δ, which, according to (8), is related to β as

$$\delta = \pi\beta.$$

In the experiments to be described below the value of δ was determined by measuring the amplitude φ some time after the free oscillations began and subsequently the amplitude φ' after the wire had made a further number of n oscillations. Thus one obtains:

$$\frac{\varphi}{\varphi'} = e^{n\delta}$$

and hence

$$\delta = \frac{1}{n} \log_e \frac{\varphi}{\varphi'}.$$

The number of oscillations $n_{0.5}$, after which the amplitude has decreased to its half-value, is given by

$$n_{0,5} = \frac{1}{\delta} \log_e 2 \approx \frac{0.7}{\delta}.$$

In our experiments δ was always between the limits 0.1 and 0.001; thus $n_{0.5}$ varied between 7 and 700 oscillations.

As already stated, β, and thus also δ, reaches its maximum value when the angular frequency ω of the oscillation and the relaxation time τ obey the relation:

$$\omega\tau = 1.$$

The level of the maximum in the curve representing the variation of δ with ω is a measure of the magnitude of the after-effect and thus a measure of the number of foreign atoms capable of jumping. Thus it is possible to learn something about this number by investigating the damping curve in the critical range.

To this end, instead of varying the frequency, the temperature of the wire may also be varied and the frequency kept constant, as will now be shown.

In order of magnitude the relaxation time τ does not differ much from the mean time that a carbon or nitrogen atom stays in a certain interstitial place, and it may therefore be represented by the formula:

$$\tau \approx \frac{a^2}{D}, \quad . \; . \; . \; . \; . \; . \; . \; . \; . \quad (12)$$

where a represents the average jump length of a carbon or nitrogen atom from one position to the other (thus in our case a is in the order of the lattice constant of α-iron), whilst D is the diffusion coefficient of carbon or nitrogen respectively in α-iron. Eq. (12) may be applied because the jumping of the C(N) atoms is a diffusion phenomenon, and for these phenomena eq. (12) is of general application.

Now D is strongly dependent upon the temperature, according to the relation

$$D = D_0 \, \mathrm{e}^{-\frac{E}{RT}},$$

in which R is the gas constant per mole and E the activation energy for an atom to pass from one interstice to another, this energy being of fairly high value, viz. about 80,000 joules (about 20,000 cal.). As a consequence τ is likewise strongly dependent upon the temperature, the relation being:

$$\tau = \tau_0 \, \mathrm{e}^{\frac{E}{RT}}.$$

To find the maximum damping it is therefore preferable to vary τ instead of ω, by changing T. When δ is plotted against $1/T$ for a constant frequency a similar curve is obtained as when δ is plotted against log $\omega\tau$ for a constant temperature. The frequency is automatically kept constant by employing the free oscillations of a torsional pendulum. If β is small, and this condition is always satisfied, the frequency of the damped oscillations is virtually the same as that of the undamped ones, and is not dependent upon time either (this had already been tacitly assumed in formula (5)).

Damping experiments

The fact that the internal friction can be used to get a better insight into the phenomena of quench ageing in iron and steel is due to the difference in behaviour of carbon or nitrogen in respect to damping, according to whether they occur in solid solution or as carbide and nitride precipitates. The reason for this is that the structure of iron carbide or iron nitride is such that there cannot be jumps of the C or N atoms in it, leading to a non-symmetrical deformation of the lattice. Any precipitation of a part of the dissolved carbon or nitrogen must, therefore, immediately find expression in a lowering of the maximum damping.

The measuring of the damping is a more sensitive and more absolute method of studying precipitation phenomena than the measuring of the hardness, since the latter depends also greatly upon the dimensions of the particles precipitated. For one and the same degree of hardness there may, therefore, be greatly different quantities of dissolved C or N. On the other hand the height of the damping peak gives at once the quantity of dissolved carbon or nitrogen, it having been found [3]) that there is a simple relationship between the two. It happens to be that the amount in solution, in weight %, corresponds fairly well with the value of β.

Before proceeding to discuss the experiments carried out to check the results of the hardness tests, and in order to gain a deeper insight into the underlying atomic processes, it should be recalled that the most important fact revealed by the hardness measurements (see I) was that the presence of 0.5% manganese in iron appreciably delays the precipitation of nitrogen but has no noticeable effect upon the precipitation of carbon.

The latter fact was immediately confirmed by damping experiments with wires made of the alloys Fe+0.04% C and Fe+0.5% Mn+0.04% C. The diameter and length of the wires and the moment of inertia of the body suspended from them were such that the periodic time of the oscillations amounted to approximately 1 second. As shown in *fig.* 7 [4]), the measurements of the damping as a function of temperature teach that immediately after quenching from 720 °C (the curves *a*) a sharp maximum in the damping occurs at about 39 °C. For Fe + 0.5% Mn + 0.04% C (dotted curve *a*) the maximum lies at the same temperature as that for Fe + 0.04% C (fully-drawn curve *a*). Also the width of the damping peak is the same in both cases. From this it is established that the rate of diffusion of carbon in iron is not influenced by the presence of manganese. Precipitation was investigated in the case of the two alloys mentioned by heating the wires for 1 hour at 100 °C after the damping measurements just described had been taken. The curves *b* show the values of δ/π after this heating. Finally the curves *c* show the damping after the wires had been heated at 100 °C for 3 hours in all. The results show that 0.5% Mn does not appreciably affect the rate of precipitation of carbon.

---

[3]) L. J. Dijkstra, Philips Res. Rep. **2**, 357-381, 1947.

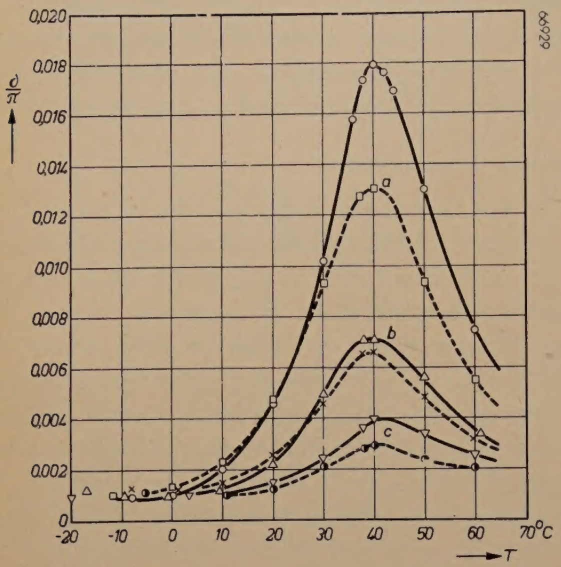


Fig. 7. The internal damping of α-iron with 0.04% carbon (fully drawn curves) and of iron with 0.5% manganese and 0.04% carbon (dotted curves) as functions of temperature. Plotted along the ordinate is the logarithmic decrement δ of the oscillations divided by π, thus the loss angle β. *a*) Immediately after quenching from 720 °C; *b*) after heating for 1 hour at 100 °C subsequent to quenching; *c*) after heating for 3 hours at 100 °C. There is considerable precipitation, which is very little affected by manganese.

Whereas 0.5% Mn has no influence upon the behaviour of carbon in iron, it appears to have a very great effect upon the behaviour of nitrogen (again in conformity with the hardness measurements). This is demonstrated in *fig. 8*, the left half of which relates to the alloy Fe + 0.02% N and the right half to the alloy Fe + 0.5% Mn + 0.02% N. Heating for three hours at 100 °C leads, according to the left-hand curves, to precipitation of more than half of the dissolved nitrogen in the alloy without manganese, but according to the right-hand curves there is hardly any precipitation in the alloy containing manganese. In addition, fig. 8 reveals some other important details as regards the position and shape of the damping curve. The presence of 0.5% Mn appears to shift its maximum from 25 °C to 32 °C, whilst the half-value width of the peak appears to have increased from 24 °C to 36 °C (ignoring the small secondary maximum at 10 °C).

The fact that after the addition of manganese the damping curve still has a peak indicates that the nitrogen atoms are still able to move to and fro between the x, y and z positions. The widening of the peak, however, signifies that there is now more than one relaxation time playing a part in this diffusion process. The shifting of the maximum from 25 °C to 32 °C further shows that after the addition of manganese the value of $\omega\tau = 1$ is reached only at a temperature which is 7 °C higher than that before the addition of Mn. For the same temperature, therefore, the mean relaxation time and thus also the mean time that the atoms stay in an x, y or z position are increased. All these facts lead to the following atomic picture.

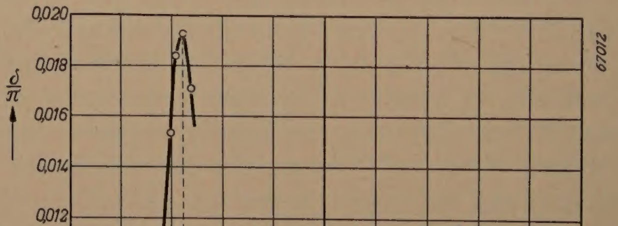


Fig. 8. The internal damping of iron with 0.02% nitrogen (left) and of iron with 0.5% manganese and 0.02% nitrogen (right) as fuctions of temperature. Plotted along the ordinate is $\delta/\pi = \beta$. *a*) Immediately after quenching from 580 °C; *b*) after 3 hours' heating at 100 °C subsequent to quenching. The alloy containing manganese shows practically no precipitation and the peak of the damping curve is displaced and broadened. The positions of the maxima are indicated by the vertical dotted lines.

---

[4]) For the sake of simplicity, in this and the two following diagrams the damping has been plotted direct as a function of the temperature in °C, though, in accordance with the foregoing, it is perhaps more correct to take as abscissa the reciprocal value of the absolute temperature. As ordinate the quantity $\delta/\pi = \gamma/\omega = \beta$ has been plotted.

For the nitrogen atoms it is more favourable, as regards energy, to occupy an interstice where they have a manganese atom as immediate neighbour ("Mn site") than an interstice which is surrounded exclusively by iron atoms ("Fe site"). The fact that the interstices are no longer energetically equivalent naturally leads to the occurrence of different relaxation times. According to the results of the experiments the nitrogen atoms are able to make jumps around the manganese atoms but they have difficulties in leaving those atoms. The precipitation of nitride does not take place at any appreciable rate, because each nitrogen atom that manages to liberate itself from a manganese atom is very soon "captured" again by the same or some other manganese atom. At any temperature there will be a certain equilibrium between occupied Fe sites and occupied Mn sites. The possible explanation of this behaviour of manganese in nitrogen-containing iron has already been dealt with in article I, so that it suffices here to recall that the cause is to be sought in the comparatively great affinity between manganese and nitrogen.

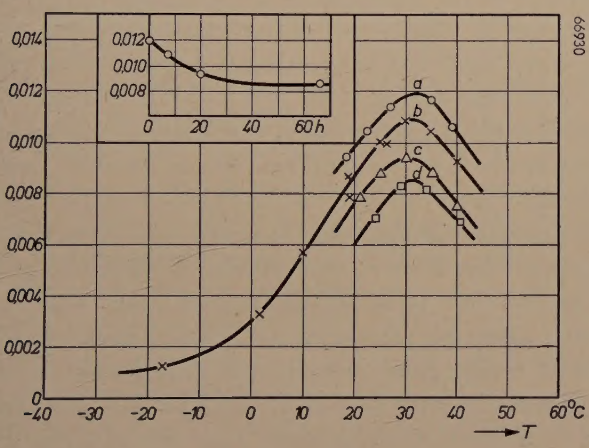


Fig. 9. Iron with 0.5% manganese and 0.02% nitrogen shows precipitation of the nitrogen (in the form of a metastable nitride) when, after quenching, the metal is heated for a long time. *a*) Damping curve immediately after quenching; *b*) after 7 hours' heating at 200 °C; *c*) after 20 hours' heating at 200 °C; *d*) after 66 hours' heating at 200 °C. In the inset, the lowering of the maximum of the damping curve due to heat treatment.

Finally, *fig. 9* shows that protracted heating at 200 °C eventually leads to the precipitation of a large part (about one-third) of the nitrogen. Even after 66 hours, however, the only precipitation is in the form of a metastable nitride, probably the phase $Fe_{16}N_2$ (also known from K. H. Jack's work; cf. footnote ^{12}) in I). This nitride and the more stable nitride Fe_4N are formed independently of each other and with very different rates of nucleation. The difficulties in nucleation for Fe_4N are so great that after 66 hours' heating at 200 °C this phase is still not formed in any perceptible quantity.

The results of the damping experiments form on the one hand a good confirmation of the conclusions drawn from the hardness tests, as set forth in I, and on the other hand lead to a deeper insight into the atomic processes underlying the phenomena observed. They also furnish an explanation of the technically very important fact that the precipitation of nitrogen in pure iron takes place at a much higher rate than in the commercial kinds of iron and steel (which always contain manganese!). Phenomena due to quench ageing may therefore arise in these types of steel even after comparatively slow cooling, since this "slow" cooling does after all take place quickly compared with the rate of precipitation.

Among these phenomena there is not only the gradual increase of the mechanical hardness but also that of the magnetic hardness of commercial mild steel. From the work of various investigators, particularly that of Köster ^{5}), it was already known that mild steel (also slowly cooled mild steel) need contain only 0.005% nitrogen for its coercive force to be doubled when heated to 100 °C for several hundred hours. This increase in coercive force is due to the precipitation of the nitrogen in the form of nitride. From the experiments described in the foregoing it may be concluded that the phenomenon of magnetic ageing is to be ascribed not to the nitrogen alone but to the simultaneous occurrence of manganese and nitrogen in the steel.

Summary. Damping experiments, at different temperatures, with a torsional wire pendulum made of iron or steel containing a little carbon or nitrogen in solution (with or without the admixture of manganese) throw some light upon the extent to which the dissolved atoms precipitate after quenching of the metal. It is shown how the internal friction, resulting from the solute atoms jumping from one interstice in the crystal lattice to another, causes the amplitude of the torsional oscillations to decay exponentially. At a certain frequency the degree of damping depends upon the rate of diffusion of the atoms and thus upon the temperature, the logarithmic decrement as a function of the temperature showing a maximum at a very definite temperature. The height of this maximum is directly proportional to the amount of dissolved atoms and thus decreases with increasing precipitation. In conformity with the conclusions drawn from hardness measurements described in the preceding article, manganese appears to have no influence upon the precipitation of carbon in carbon-containing iron after quenching, whilst the precipitation of nitrogen in nitrogen-containing iron is greatly retarded by manganese, so much so that even after 66 hours' heating at 200 °C (after quenching) only about one-third of the nitrogen is precipitated in the form of a (metastable) nitride. The results of the damping experiments throw light upon the atomic processes responsible for this behaviour and make it possible to give a new explanation of the mysterious phenomenon of "magnetic ageing".

---

^{5}) W. Köster, Z. anorg. allg. Chem. **179**, 297-308, 1929.

BOOK REVIEW

Application of the electronic valve in radio receivers and amplifiers. A.F. amplification, output amplification and power supply, by B. G. Dammers, J. Haantjes, J. Otte and H. van Suchtelen, 452 pages, 343 illustrations. — Edited by Philips Technical Library (Book 5 of the series on electronic valves) publishers N.V. Meulenhoff & Co., Amsterdam, 1951.

The first part (book 4) of the above-mentioned work [1]) dealt with R.F. and I.F. amplification, mixing and detection. The second part now published deals with A.F. voltage amplification, output stages and supply. In a third part now in course of preparation other important subjects related to the application of the electronic valve in receivers and amplifiers will be discussed.

The second part is a worthy continuation of the first, the authors placing their vast experience at the disposal of everyone having anything at all to do with receivers and amplifiers. Practically every aspect of the application of the electronic valve is thoroughly dealt with in all its details as far as it is related to the subjects treated. Numerous calculations are given by way of illustrating the subject matter.

The book is divided into three chapters: A.F. Amplification, the Output Stage and the Power Supply. The first chapter deals in succession with A.F. amplifying circuits, phase-inverting stages, the response curve, calculation of the A.F. transformer and non-linear distortion. Under the Output Stage are discussed: the various adjustments (class A, class B, class AB, comparison of these adjustments and double-tone testing), distortion, behaviour under complex load, deviation from normal valve adjustments and overload phenomena. The third chapter deals with the filament-current supply, rectifying circuits, calculation of H.T. rectifiers and circuits for stabilizing supply voltages. For the benefit of readers desiring to go more deeply into certain subjects a bibliography is appended at the end of each chapter.

This book will prove to be an excellent vademecum for many. The uninitiated would never suspect that so much could be said about the subjects dealt with. Particularly the chapter on the output stage goes to great length, covering more than half the book and dealing with the subject in a manner unequalled elsewhere. Special attention is drawn to what is written about the double-tone method.

The clear and thorough way of treatment makes the book pleasant to read. This thoroughness might have its shadow side for those who are not yet familiar with the subject, in that they might not be able to see the wood for the trees, but the logical manner in which the subject matter is set forth ensures that the main thread need not be lost.

Notwithstanding this extensive treatment the subject of inverse feedback, which is of such great importance with A.F. amplifiers, has not been discussed in this book but kept for the third part (book 6). It might be thought that it would have been better to keep together all the problems arising in A.F. amplification, so as to facilitate reference. But with such thorough treatment as is now presented, this book would have become too voluminous if the subject of inverse feedback had been also included. It is therefore with interest that we look forward to the appearance of the third part, as an essential supplement to part 2.

TH. J. WEIJERS

---

[1]) Discussed in vol. **12** of this journal, page 184, 1950.